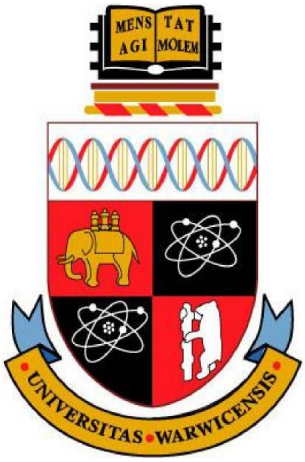
Investigations of Carbocyclisation Reactions in Prodiginine Alkaloid Biosynthesis

Rebin M. Salih

Supervisor: Prof. Gregory L. Challis

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Chemistry



Department of Chemistry, University of Warwick

December 2017

Table of Content

Acknowledgements

Firstly, I would like to express my thanks to my supervisor, Professor Gregory L. Challis for giving me the opportunity to work on this PhD project under his guidance, encouragement and supervision. His creativity, patience and constructive comments were always invaluable to this work. I would like to extend my gratitude to my advisory panel members, Dr. Christophe Corre and Dr. Andrew Marsh for their helpful discussions throughout my work.

Secondly, I would like to thank Dr. Lijiang Song for his help, to Dr. Paulina K. Sydor for her helpful guidance and Dr. Lona Alkhalaf for her valuable comments and advice on my thesis. For their technical advice in the lab, I wish to thank all the members of Challis group (past and present); Dr. David Withall, Dr. Lauren Ray, Dr. Chuan Huang, Dr. Daniel Zabala-Alvarez, Dr. Yuki Inahashi, Dr. Yousef Dashti, Dr. Emzo De Los Santos, Dr. Matthew Jenner, Dr. Douglas Roberts, Dr. Jade Ronan, Dr. Daniel Griffiths, Dr. Joleen Masschelein, Dr. Simone Kosol, Dr. Vincent Poon, Dr. Matias Rey, Dr. John D. Sidda, Dr. Gideon Idowu, Dr. Shanshan Zhou, Dr. Joshua Cartwright, Chris Perry, Marianne Costa, Ruby Awodi, Richard Gibson, Rakesh Saroay, Xinyun Jian, Christian Hobson, Dr. Goran Rashid and Dr. Rahman Rahmanpur (Bugg group), our technicians Anne Smith and Dr. Magda Moses, for their friendly advice and all of the other people in the Chemical Biology Research Facility.

Finally, I would like to thank my beloved wife (Nian) who was always looking after our son (Nishan), mother, sisters and brothers for their endless support and encouragement through my PhD.The thesis is dedicated to the memory of my late father.

Thanks to the HCDP-Program of Kurdistan Region-Iraq for the sponsorship.

Declaration

The experimental work reported in this thesis is original research carried out by the author, unless otherwise stated, in the Department of Chemistry, University of Warwick between November 2012 and March 2017. No material has been submitted for any other degree, or at any other institution.

Results from authors are referenced in the usual manner throughout the text.

_____ Date: _____

Rebin M. Salih

Abstract

Prodiginine alkaloids, including streptorubin B, metacycloprodigiosin and roseophilin, are a family of red-pigmented pyrrole-containing specialised metabolites with immunosuppressive, anticancer and antimalarial properties. The final steps in streptorubin B and metacycloprodigiosn biosynthesis are oxidative carbocyclisations catalysed by Rieske oxygenases RedG and McpG respectively. These enzymes utilize the same substrate (undecylprodigiosin) to give cyclised products with different regio and stereochemistry.

In order to further investigate the McpG catalysed reaction and to develop a system that could be used to probe the mechanism, *mcpG* was expressed in *S. albus* and *S. coelicolor* M511. Cyclic compound was observed following feeding of undecylprodigiosin to *S. albus/mcpG*. Since production was low, the role of McpH in forming a complex with McpG was investigated. Unfortunately, whilst McpH was shown to catalyse the condensation of 2-undeyclpyrrole (2-UP) and 4-methoxy-2, 2' bipyrrole-5-carboxaldehyde (MBC), co-expression of *mcpH* with *mcpG* did not improve production of metacycloprodigiosin. Since whole genome sequencing showed that the *mcpG* expression construct originally used to prove the function of this gene contained a 5' truncation due to a sequencing error, full length *mcpG* was expressed in a *redG* mutant of *S. coelicolor*. This resulted in production of metacyloprodigiosin, which was confirmed by mass spectrometry and ^{1}H and ^{13}C NMR spectroscopy.

The *rph* gene cluster reported to be responsible for the biosynthesis of prodigiosin R1 and roseophilin in *S. griseoviridis* contains four *redG* orthologues (*rphG*, *rphG2*, *rphG3* and *rphG4*). Unfortunately, neither expression of the *rphG* genes in the *redG* mutant of

S. coelicolor, nor the feeding of 11-methyldodecylprodigiosin to *S. albus* separately expressing each of the *rphG* genes resulted in the production of carbocylic products.

Bioinformatics analysis of the *S. longispororuber* genome sequence revealed several cryptic natural product biosynthetic gene clusters. The likely products of some of these clusters are discussed in the light of predictive sequence analysis.

Abbreviations

A	Adenylation (domain)
a. a.	amino acid
ACP	ACP Acyl Carrier Protein
AHFCAs	2-Alkyl-4-hydroxylmythylfuran-3-carboxylic acid
Amp	Ampicillin
AT	Acyl Transferase (domain)
ATP	Adenosine triphosphate
AROs	Aromatases
BC	Before Christ
BLAST	Basic Local Alignment Search Tool
Bp	Base pairs
C	Condensation (domain)
CDA	Calcium-dependent antibiotic
CoA	Coenzyme A
CYCS	Cyclases
DH	DH Dehydratase (domain)
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
E	Epimerisation (domain)
ECH	Enoyl-CoA hydratase
EDTA	Ethylenediaminetetraacetic acid
EIC	Extracted ion chromatogram
ER	Enoyl Reductase (domain)
HMGS	3-hydroxyl-3-methylglutaryl-CoA synthase
FAS	Fatty Acid Synthase
GBLs	gamma-butyrolactones
HPLC	High-performance liquid chromatography
Hyg	Hygromycin
ID	Identity
Kan	Kanamycin
Kb	kilo base pairs
KASII	β-ketoacyl-ACP synthase II
KASIII	β-ketoacyl-ACP synthase III
KR	Ketoreductase (domain)
KS	Ketosynthase (domain)
LB	Luria-Bertani (Medium)
LC-MS	Liquid Chromatography – Mass Spectrometry
M	Methylation (domain)
MMFs	methylenomycin Furans
MBC	4-methoxy-2,2 -bipyrrole-5-carboxaldehyde
min.	Minute
MS	Mass Spectroscopy
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
NCBI	National Centre for Biotechnology Information
NDO	Naphthalene dioxygenase
NIR-MCD	Near Infrared Magnetic circular Dichroism

NMR	Nuclear Magnetic Resonance
NRPS	Nonribosomal peptide synthetase
Nt	Nucleotide
OAS	Oxoamine synthase
OD	Optical Density
ORF	Open Reading Frame
oriT	Origin of Transfer
PCR	Polymerase Chain Reaction
PCP	Peptidyl Carrier Protein
PEPS	phosphoenol pyruvate synthase
PKS	Polyketide synthase
PLP	Pyridoxal phosphate
PPDK	Phosphate Pyruvate Dikinase
PPTase	Phosphopantethienyl transferase
RBS	Ribosome binding site
rpm	revolutions per minute
SAM	S-adenosyl methionine
SFM	Soya Flour Mannitol medium
SMM	Supplemented Minimal Medium
T	Thiolation (domain)
Taq	*Thermus aquaticus* (polymerase)
TBE	Tris-boric acid EDTA buffer
TE	Thioesterase (domain)
Tet	Tetracycline
Tris	Tris(hydroxymethyl)aminomethane
2-UP	2-Undecylpyrrole

List of Figures

highlighted in blue and green respectively; a mutation from aspartate in NDO to glutamate in RedG and McpG is highlighted in red.

Figure 1.13 Reaction catalysed by naphthalene dioxygenase (left), proposed [2Fe-2S] Rieske cluster (blue), non-haem iron active site (green) and bridging aspartate residue (red).

Figure 1.14 Proposed mechanism for the RedG-catalised oxidative carbocyclisation of undecylprodigiosin (**2**) to streptorubin B (**3**).

Figure 1.15 Regio-stereodivergent oxidative carbocyclisation reactions catalysed by RedG and McpG to form streptorubin B (**3**) and metacycloprodigiosin (**19**).

Figure 1.16 proposed functions of RphG, RphG2, RphG3 and RphG4 in *Streptomyces griseoviridis*.

Figure 2.1 Metacycloprodigiosin production by *S. longispororuber*. A. EIC for *m/z* range 392-394 (corresponding to $[M+H]^+$ for undecylprodigiosin (**2**) and metacycloprodigiosin (**19**) from LC-MS analysis of acidified organic extract of *S. longispororuber*. B. HR-MS confirming the molecular formula of **19** (calculated for $C_{25}H_{35}N_3O$:392.2696, found: 392.2703) and for **2** (calculated for $C25H_{37}N_3O$: 394.2856, found: 394.2869).

Figure 2.2 The pOSV556 vector.

Figure 2.3 Agarose gel electrophoretic analysis of *Xho*I restriction digests of pOSV556*redG* (lane 1) and pOSV556*mcpG* (lane 2). The lane marked M contains molecular size markers.

Figure 2.4 Activity assay of McpG and RedG in *S. coelicolor* M511. EIC for *m/z* range 392-394 (corresponding to $[M+H]^+$ for undecylprodigiosin (**2**) and metacycloprodigiosin (**19**) from LC-MS analysis of acidified organic extracts of *S. coelicolor* M511 (top trace), *S. coelicolor* M511/pOSV556*redG* (middle trace) and *S.*

coelicolor M511/pOSV556*mcpG* (bottom trace) fed with undecylprodigiosin extracted from *S. coelicolor* W31(Δ*redG*).

Figure 2.5 Activity assay of McpG and RedG in *S. albus*. EIC for *m/z* range 392-394 (corresponding to $[M+H]^+$ for undecylprodigiosin (**2**), streptorubin B (**3**) and metacycloprodigiosin (**19**) from LC-MS analysis of acidified organic extracts of *S. albus* (top trace), *S. albus*/pOSV556*redG* (middle trace) and *S. albus*/pOSV556*mcpG* (bottom trace) fed with undecylprodigiosin extracted from *S. coelicolor* W31(Δ*redG*).

Figure 2.6 Agarose gel electrophoretic analysis of pOSV556 (lane 1), pOSV556*mcpH* (lane 2) and pOSV556*mcpHG* (lane 3) digested with *Bam*HI. The lane marked M contains1kb molecular size markers.

Figure 2.7 Agarose gel electrophoretic analysis of A. PCR amplification of *mcpH*. B. Digestion of pOSV556*mcpH* with *Hin*dIII/*Stu*I and *Xho*I/*Pst*I. The lanes marked M contains1kb molecular size markers Expected band sizes are labelled.

Figure 2.8 Activity assay for McpH in *S. albus*. A. EIC for *m/z* range 392-394 (corresponding to $[M+H]^+$ for undecylprodigiosin (**2**) and metacycloprodigiosin (**19**) from LC-MS analysis of acidified organic extracts of *S. longispororuber* (top trace), *S. albus* (middle trace) and *S. albus*/pOSV556*mcpH* (bottom trace) fed with 2-UP (**15**) and MBC (**16**). B. HR-MS confirming the molecular formula of **2** (calculated for $C_{25}H_{37}N_3O$: 394.2856, found; 394.2856).

Figure 2.9 Agarose gel electrophoretic analysis, A. PCR of *mcpHG* (3927bp). B. Restriction digestion of *mcpHG* PCR products with *Xho*I, C. Restriction digestion of pOSV556 (Control), pOSV556*mcpHG* clone no.3, (2) and pOSV556*mcpHG* clone no.8 with *Hin*dIII and *Stu*I, *Xho*I. The lane marked M contains 1kb molecular markers. Expected band sizes are labelled in white.

Figure 2.10 Activity assay of McpHG in *S. albus*. A. EIC for *m/z* range 392-394 (corresponding to $[M+H]^+$ for undecylprodigiosin (**2**) and metacycloprodigiosin (**19**) from LC-MS analysis of acidified organic extracts of *S. albus* (top trace), *S. albus*/pOSV556*mcpH* (middle trace) and *S. albus*/pOSV556*mcpHG* (bottom trace) fed with 2-UP (**15**) and MBC (**16**). B. HR-MS confirming the molecular formula of **19** (calculated for $C_{25}H_{35}N_3O$: 392.2696, found: 392.2698) and for 2 (calculated for $C_{25}H_{37}N_3O$: 394.2856, found: 394.2860).

Figure 2.11 Protein sequence alignment of the McpG and McpG (updated) from *S. longispororuber*.

Figure 2.13 Activity assay of McpH in *S. albus*. A. EIC for *m/z* range 392-394 (corresponding to $[M+H]^+$ for undecylprodigiosin (**2**) and metacycloprodigiosin (**19**) from LC-MS analysis of acidified organic extracts of *S. longispororuber* (top trace), *S. albus* (middle trace) and *S. albus*/pOSV556*mcpHG* (bottom trace) fed with 2-UP (**15**) and MBC (**16**). B. HR-MS confirming the molecular formula of **19** (calculated for $C_{25}H_{35}N_3O$: 392.2696, found: 392.2685) and for 2 (calculated for $C_{25}H_{37}N_3O$: 394.2856, found: 394.2843).

Figure 2.14 Agarose gel electrophoretic analysis, A. PCR amplification of *mcpG*. B. pOSV556*mcpG* digested with *Hin*dIII and *Xho*I. The lane marked M contains 1kb molecular markers and the HR contains high range molecular markers.

Figure 2.15 Metacycloprodigiosin production by *S. longispororuber*. A. EIC for *m/z* range 392-394 (corresponding to $[M+H]^+$ for undecylprodigiosin (**2**) and metacycloprodigiosin (**19**) from LC-MS analysis of acidified organic extract of *S. longispororuber* (top trace), *S. coelicolor* W31 (middle trace) and *S. coelicolor* W31/pOSV556*mcpG*. B. HR-MS confirming the molecular formula of **19** (calculated

for $C_{25}H_{35}N_3O$: 392.2696, found: 392.2695) and **2** (calculated for $C_{25}H_{37}N_3O$: 394.2856, found: 394.2869).

Figure 2.16 NMR analysis of metacycloprodigiosin (**19**).

Figure 2.17 A. ^{1}H NMR spectra in $CDCl_3$ (700 MHz) of metacycloprodigiosin. The characteristic signals at 0.2 ppm for metacycloprodigiosin (**19**) is highlighted by a box, a- methane protons, b- hydrocarbon protons α to the pyrrole ring. B. ^{13}C NMR spectra ($CDCl_3$, 175 MHz) of metacycloprodigiosin (**19**).

Figure 3.1 A. *rph* cluster proposed to encode prodiginine biosynthesis in *Streptomyces griseoviridis.* B. Three C-C bond forming steps required for the biosynthesis of roseophilin and prodigiosin R1.

Figure 3.2 Protein sequence alignment of RedG and RphGs 1, 2, 3 and 4. The conserved residues highlighted in blue ligate the [2Fe-2S] cluster (missing in the RphG3) and Fe(II) atom, (A Glu residue in RedG and RphGs) highlighted in red is proposed to perform electron transfer between the [2Fe-2S] and the Fe(II) atom.

Figure 3.3 A. Agarose gel electrophoretic analysis of restriction digestion of pOSV556*rphG1*, pOSV556*rphG2*, pOSV556*rphG3* and pOSV556*rphG4* with *Bam*HI, 1146bp, 1329bp, 981bp, and 1101bp, respectively. The lane marked M contains 1kb molecular size. B and C. PCR amplification of *rphG1*, *rphG2*, and *rphG4* from *S. albus* transconjugants. The lane marked M contains1kb molecular size markers. PCR amplification of *rphG3*. The lane marked M contains1kb molecular size markers. P is pOSV556 and C is control.

Figure 3.4 Agarose gel electrophoretic analysis of A. *rphG*3 PCR product, B. Restriction digestion of pOSV556 (vector) and pOSV556*rphG*3 clone no.2 and clone no.3. C**.** *rphG*3 PCR from *S. albus* transconjugants (981bp). The lane marked M contains 1kb molecular markers and the MR contains middle range molecular markers.

Figure 3.5 Activity of RphGs in *S. albus*. EIC for *m/z* range 392-422 (corresponding to $[M+H]^+$ for 11-methyl-dodecylprodigiosin (*m/z* = 422.31) from LC-MS analysis of acidified organic extract of A. *S. albus*. B. *S. albus*/pOSV556*rphG*1. B. *S. albus*/pOSV556*rphG2*. C. *S. albus*/pOSV556*rphG*3. D. *S. albus*/ pOSV556*rphG*4 (top traces of A, B, C and D are *S. albus*) fed with synthetic 11-methyl-dodecylprodigiosin (*m/z* = 422.3039).

Figure 3.6 Activity of RphGs in *S. coelicolor* W31Δ*redG*. A. EIC for *m/z* = 392-394 (corresponding to $[M+H]^+$ for undecylprodigiosin (**2**) and metacycloprodigiosin (**19**)/streptorubin B (**3**) from LC-MS analysis of acidified organic extracts of 1. *S. coelicolor* M511, 2. *S. coelicolor* W31 (Δ*redG*), 3. *S. coelicolor* W31/pOSV556*rphG*1, 4. *S. coelicolor* W31/pOSV556*rphG*2 and 5. *S. coelicolor* W31/pOSV556*rphG*4. B. HR-MS confirming the molecular formula of **2** (calculated for $C_{25}H_{37}N_3O$ = 394.2856, found 394.2777).

Figure 3.7 Cloning of synthetic *rphH* gene into previously prepared pOSV556 containing *rphG1*, *rphG*2, *rphG*3 or *rphG*4 separately.

Figure 3.8 Agarose gel electrophoretic analysis, A. PCR product of *rphH*. B. Digestion of synthetic *rphH* cloned into pOSV556 with *Pml*I and *Hin*dIII. C. Digestion of sub-cloned synthetic *rphH* gene into previously recombinant pOSV556 holding *rphG*1, *rphG*2, *rphG*3 and *rphG*4 separately, with *Pml*I and *Xho*I. The lane marked M contains 1kb molecular markers and the MR contains middle range molecular markers.

Figure 3.9 Activity of RphH from *S. griseoviridis*. A. EIC for *m/z* = 422 (corresponding to $[M+H]^+$ for 11-methyldodecylprodigiosin (**26**) from LC-MS analysis of acidified organic extract of *S. albus* (top trace), *S. albus*/pOSV556*mcpH* (middle trace) and *S. albus*/pOSV556*rphH* (bottom trace) fed with 11-methyldodecylpyrrole and

MBC. B. HR-MS confirming the molecular formula of **26** (calculated for $C_{27}H_{39}N_3O$= 422.3039, found; 422.3026).

Figure 3.10 A) Protein sequence alignment of the RedH, McpH and RphH, The two places where the different RphH variants begin are marked in red. B) DNA sequence of *rphH*, the two alternative start codons are marked in green.

Figure 3.11 Agarose gel electrophoretic analysis, A. Single digestion of 12 clones of pOSV556*rphH* variant 1 and pOSV556*rphH* variant 2. B. PCR product of *rphH* variant 1 and variant 2 digested with *Kpn*I. C. pOSV556*rphH* variant 1 and pOSV556*rphH* variant 2 digested with *Pml*I and *Hin*dIII.

Figure 3.12 Activity of RphH from *S. griseoviridis*. A. EIC for *m/z* = 422 (corresponding to $[M+H]^+$ for (**26**) from LC-MS analysis of acidified organic extract of *S. albus/*pOSV556*rphH* (top), *S. albus*/pOSV556*rphH* variant 1 (second from top) *S. albus*/pOSV556*rphH* variant 2 (second from bottom) and *S. albus*/pOSV556*mcpH* (bottom) fed with 11-methyldodecylpyrrole and MBC. B. HR-MS confirming the molecular formula of **26** (calculated for $C_{27}H_{39}N_3O$= 422.3039, found; 422.3027).

Figure 3.13 Activity of RphG from *S. griseoviridis*. A. EIC for *m/z* =422 (corresponding to $[M+H]^+$ for undecylprodigiosin (**2**) and metacycloprodigiosin (**19**)/streptorubin B (**3**) from LC-MS analysis of acidified organic extract of *S. coelicolor* M511 (top), *S. coelicolor* Δ*redG* and (second from top), *S. coelicolor* Δ*redG* pOSV556*rphG* with natural RBS (second from bottom) and *S. coelicolor* Δ*redG* pOSV556*rphG* with universal RBS (bottom). B. HR-MS confirming the molecular formula of **2** (calculated for $C_{25}H_{37}N_3O$= 394.2856, found; 394.2799).

Figure 3.14 Protein sequence alignment of the *rphG1* (2009) and *rphG1* updated (2016) showing 59 missing amino acids in the C-terminal of the protein.

Figure 3.15 Activity of RphG from *S. griseoviridis*. A. EIC for *m/z* =422 (corresponding to $[M+H]^+$ for undecylprodigiosin (**2**) and metacycloprodigiosin (**19**)/streptorubin B (**3**) from LC-MS analysis of acidified organic extract of of *S. coelicolor* M511 (top trace), *S. coelicolor* Δ*redG* and (middle trace) and Δ*redG* pOSV556*rphG* (bottom trace). B. HR-MS confirming the molecular formula of **2** (calculated for $C_{25}H_{37}N_3O$= 394.2856, found; 394.2871).

Figure 4.1 Cluster cluster alignment; *mcp* cluster from *S. longispororuber* (upper) and red cluster from *S. coelicolor* (bottom).

Figure 4.2 Protein sequence alignment of the McpH and McpH (updated) from *S. longispororuber*.

Figure 4.3 Protein sequence alignment of the McpG and McpG (updated) from *S. longispororuber*.

Figure 4.4 Location of gene clusters on the chromosome of *S. longispororuber* as predicted by antiSMASH; five clusters that showed 100% similarity (in black-bold) and two clusters appeared to be intreasting; cluster 15 at 3.448856 nt and cluster 24 at 4.755119 nt (in red).

Figure 4.5 (Upper). Organisation of gene clusters from *S. longispororuber* that showed %100 similarity; A. Gene cluster 5, albaflavenone. B. Gene cluster 20, pamamycin. C. Gene cluster 27, 2-methyl-isoborneol. D. Gene cluster 30, coelichelin. E. Gene cluster 32, ectoine. F. Gene cluster 13, bacterioaminohopanetriol (gene 10 which is of unknown function is absent). G. Gene cluster 35, desferrioxamine. (Bottom). Structure of the metabolites encoded by the gene clusters.

Figure 4.6 A. Muribactin biosynthetic gene cluster from *Actinosynnema mirum*. B. Muribactin biosynthetic gene cluster from *S. longispororuber* (cluster 8, 2161361–2219220 nt). C. Muribactin structure showing the acyl hydroxamic acid ester group.

List of Tables

Introduction

1.1 *Streptomyces coelicolor* A3(2) as a model producer of specialised metabolites

Streptomyces is the largest genus of Actinobacteria with more than 500 species currently known (Kieser *et al*., 2000). *Streptomyces* reside in soil and water and have a complex life cycle. Actinobacteria play a major role in decomposition of organic matter in soil contributing in part to the earthy odour of soil, which results from production of a volatile metabolite, geosmin (Guest *et al*., 2003).

In addition, Actinobacteria produce a wide range of structurally diverse specialised metabolites that include over two-thirds of the clinically useful antibiotics, of which 80% are produced by *Streptomyces* ssp. Many of these compounds have pharmaceutical application in human medicine as antibacterial, antitumor, antifungal and immunosuppressant agents (Kieser *et al*., 2000). Alternatively, some actinomycetes are pathogenic to plants, animals and human such as *Mycobacterium tuberculosis* which causes tuberculosis (Madigan, and Martinko, 2005).

In 2002, the complete genome sequence was published (Bentley *et al*., 2002). The linear chromosome is 8.7 Mb long with 7,800 predicted protein encoding genes. As well as the linear chromosome, *S. coelicolor* A3(2) contains a 360 kb linear plasmid called SCP1 (Bentley *et al*., 2004) and a 30 kb circular plasmid called SCP2 (Haug *et al*., 2003). Genome sequence analysis showed that over 20 gene clusters direct the production of known or predicted specialised metabolites (Bentley *et al*., 2002). This genetic information made *S. coelicolor* A3(2) an ideal model strain for the study of the biosynthesis of specialised metabolites such as the prodiginines (Bentley *et al*., 2002).

1.2 Specialised metabolites of *Streptomyces coelicolor* and their biological activities

Prior to the complete genome sequence of *S. coelicolor* A3(2), some of the natural product biosynthesis genes within the genome were already known to direct the production of specialised metabolites such as actinorhodins (**1**), prodiginines (**2** and **3**), methylenomycins (**4** and **5**), calcium dependent antibiotics (CDAs) (**6**), a grey spore pigment and gamma-butyrolactones (GBLs) SCB1 (**7**) (Figure 1.1).

In addition, genome sequence analysis of *S. coelicolor* A3(2) has led to the identification of many cryptic gene clusters that direct the biosynthesis of previously unknown metabolic products. These cryptic gene clusters encode a wide range of enzyme classes commonly involved in specialised metabolite biosynthesis, such as type I modular and iterative polyketide synthases (PKSs), type II PKSs, type III PKSs, nonribosomal peptide synthetases (NRPSs), NRPS-independent siderophore (NIS) synthetases and terpene synthases (Bentley *et al*., 2002).

Using a variety of approaches, the metabolic products of several of these cryptic gene clusters, including geosmin (**8**) methyl-isoborneol (**9**), albaflavenone (**10**), coelichelin (**11**), desferrioxamine (**12**), methylenomycin furans (MMFs) (**13**), and germicidins (**14**), (Gust *et al*., 2003; Lautru *et al*., 2005; Song *et al*., 2006; Corre *et al*., 2008; Zhao *et al*., 2008) and coelimycin (**24**) (Gómez-Escribano *et al*., 2012) have since been discovered (Figure 1.1). There are different ways to regulate the production of specialised metabolites in *Streptomyces* species, including by pathway-specific transcriptional regulators, by pleiotropic mechanisms and by coordination of antibiotic production and morphological development (Bibb, 2005).

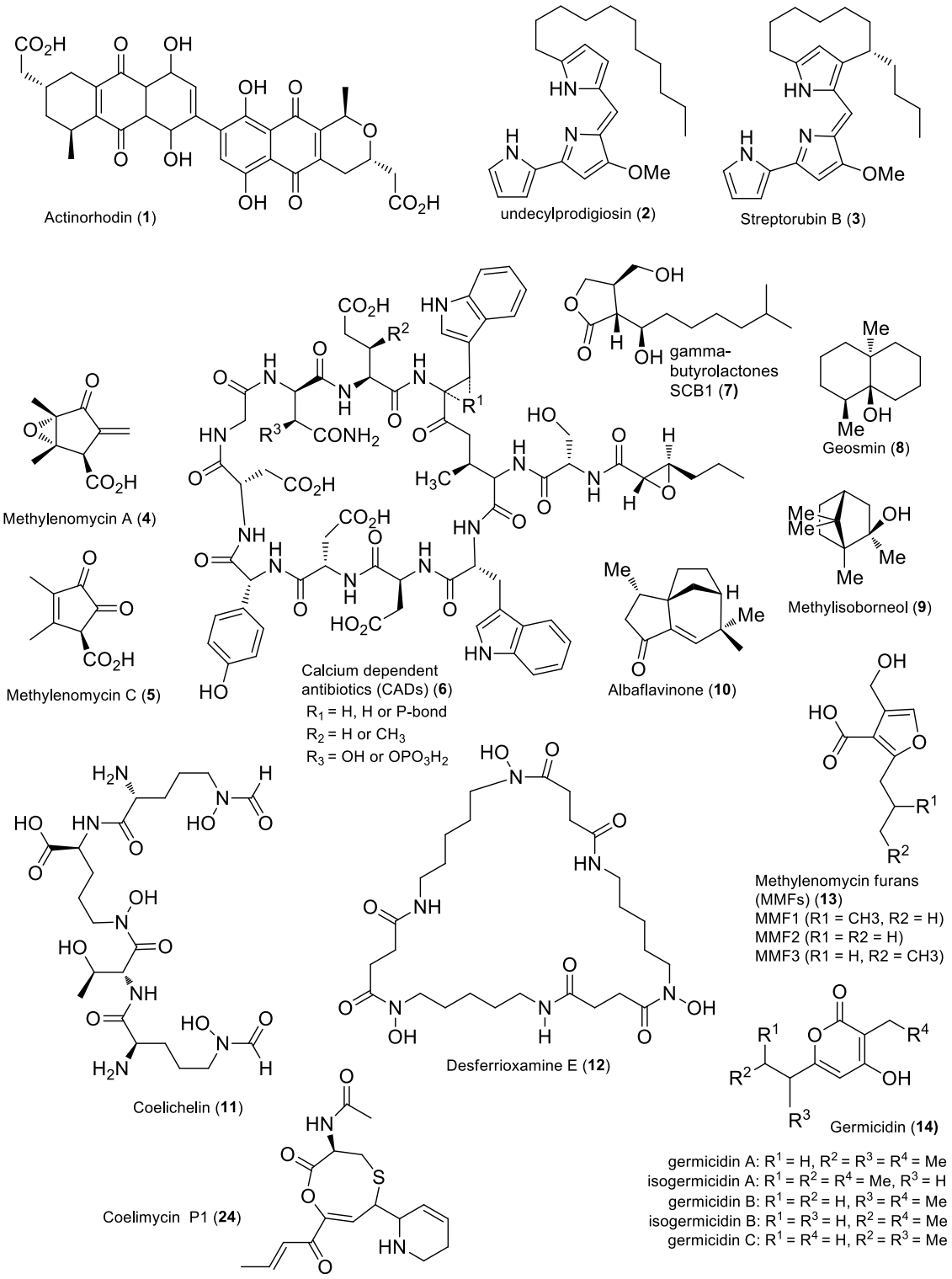


Figure 1.1 Examples of specialised metabolites produced by *Streptomyces coelicolor* A3(2).

1.2.1 Actinorhodins

Actinorhodin (**1**) (Figure 1.1), a blue-pigmented antibiotic produced by *S. coelicolor* (Wright and Hopwood, 1976) belongs to a class of aromatic polyketides, the benzoisochromanequinones (BIQS). Its blue colour in high pH and red colour in low pH results from its dimeric structure with two naphthazarine rings connected by a C-C bond (Hopwood, 1997). Actinorhodin is the first antibiotic whose biosynthetic gene cluster, which comprises 22 genes, was cloned in its entirety and has been used as a model system for genetic and biochemical studies of bacterial aromatic type II PKSs.

Some actinorhodins possess anticancer, antibacterial, anticoccidial or platelet aggregation inhibitory properties (Brockmann *et al*., 1950). Actinorhodin biosynthesis begins with the production of an octapeptide catalysed by the minimal typeII PKSs. The octapeptide subsequently undergoes tailoring modifications such as ketoreductase (KRs), aromatases (AROs), cyclases (CYCS), oxygenases and others to yield the Actinorhodin antibiotic (Fernandez-Moreno *et al*., 1992).

1.2.2 Methylenomycin

The original isolation and structure of methylenomycin A and B were first reported in 1974 from *Streptomyces violaceoruber* (Haneishi *et al*., 1974). Methylenomycin A and its desepoxy-4, 5-dehydro derivative, methylenomycin C (**5**) were later shown to be produced by *S. coelicolor* (Wright and Hopwood, 1976; Hornemann and D.A., 1978). Methylenomycin B was then demonstrated to derive from the spontaneous degradation of methylenomycin A (**4**) in acid conditions (Figure 1.1) (Corre and Challis, 2005).

Genetic studies confirmed that the large linear plasmid SCPI which is not essential for *S. coelicolor* growth carries the *mmy* gene cluster which comprise twenty one genes

responsible for the self-resistance, regulation of production and biosynthesis of methylenomycin in *S. coelicolor* A3(2) (Kirby and Hopwood, 1977; Bentley *et al*., 2004).

Labelling experiments showed that methylenomycins derive from two acetate units and one pentulose unit which are condensed to form the common 4,4-dimethyl-3-oxo-2 methylene cyclopentane carboxylic acid core of **4** and **5** (Corre and Challis, 2005).

Methylenomycin A showed a wide spectrum of antibacterial activity against Gram-positive bacteria and some Gram-negative strains especially *Proteus* species.

1.2.3 Calcium Dependent Antibiotics (CDAs)

The calcium-dependent antibiotics (CDAs) (**6**) (Figure 1.1) produced by *S. coelicolor* are acidic lipopeptides containing several nonproteinogenic amino acids residues linked to a fatty acid chain and they require calcium ions for their antibacterial activity (Kempter *et al*., 1997). They were identified as specialised metabolic products of *S. coelicolor* in 1978 (Lakey *et al*., 1983).

CDAs belong to a family of anionic lipo-undecapeptides that also includes daptomycin (produced by *Streptomyces roseosporus*) and friulimicins (produced by *Actinoplanes friuliensis*) (Vertesy *et al*., 2000). The CDA gene cluster comprises forty genes, twenty of which encode the biosynthetic enzymes required for the production of the CDAs, three genes in the regulation of biosynthesis, four in self-resistance and thirteen with unknown function (Hojati *et al*., 2002). Three NRPSs, a fatty acid synthase and enzymes involved in precursor biosynthesis and peptide tailoring catalyse the biosynthesis of CDAs (Hojati *et al*., 2002).

CDAs are active against a wide range of Gram-positive pathogens including strains resistant to the currently used antibiotics. It has been reported that they cause membrane-depolarisation and cell death by inducing the formation of cation specific channels within the bacterial membrane (Lakey *et al*., 1983). Daptomycin, an example of a CDA, which is active against methicillin-resistant *Staphylococcus aureus* and vancomycin resistant *Enterococci*, has been approved for the treatment of skin and soft tissue infections (Juang *et al*., 2004).

1.2.4 Albaflavenone

Albaflavenone (**10**) (Figure 1.1) was shown to be produced by *S. coelicolor* A3 (2). This antibiotic is biosynthesised by a sesquiterpene synthase encoded by *sco5222* which catalyses cyclisation of farnesyl diphosphate to the novel tricyclic hydrocarbon, epi-isozizaene. A cytochrome encoded by *sco5223* then catalyses two sequential allylic oxidations to form albaflavenone (Zhao *et al*., 2008).

1.2.5 Siderophores

Siderophores are released by many microorganisms to scavenge ferric iron from the environment. Iron is an essential element for cell respiration and DNA synthesis (Wandersman and Delepelaire, 2004). Imbert *et al* have reported that under iron-deficient conditions *S. coelicolor* produces hydroxamate siderophores desferrioxamine G and E (Imbert *et al*., 1995). Using bioinformatics analysis of the *S. coelicolor* genome sequence, the production of a second type of siderophores, coelichelin was predicted (Challis and Ravel, 2000) and subsequently isolated and structurally characterised (Lautru *et al*., 2005). The new desferrioxamine produced by *S. coelicolor* lack exponent of desferrioxamine B and E (**12**) (Figure 1.1) A cluster of four genes (*desA-D*) was

identified from *S. coelicolor* to be encoding for the biosynthesis of desferrioxamine. The biosynthetic pathway of desferrioxamine belongs to a family of pathways for siderophores biosynthesis (Barona-Gomez *et al*., 2006).

The *cch* gene cluster responsible for the biosynthesis of coelichelin encodes an NRPS (CchH) containing three modules predicted to incorporate N-formyl-N-hydroxyornithine, threonine and N-hydroxyornithine to form a tripeptide (Challis and Ravel, 2000). Since coelichelin contains a tetra-peptide rather than a tri-peptide, the first module is proposed to act iteratively and incorporate two molecules of N-formyl-N-hydroxyornithine into coelichelin (Lautru *et al*., 2005).

Barona-Gomez *et al* showed that at least one of desferrioxamine B, desferrioxamine E or coelichilin is required for *S. coelicolor* growth on a xeno-siderophore free colloidal silica medium (2006).

1.2.6 Signalling molecules

Gamma-butyrolactones (GBLs) produced by actinomycetes are low molecular weight molecules that regulate antibiotic production, morphological development and other environmental signals (Takano, 2006). Investigations of signalling molecules in *S. coelicolor* showed that at least seven GBLs are produced by this strain, SCB1 (**7**) (Figure 1.1), the most well studied is proposed to control the expression of a pathway-specific regulatory gene in the type I PKS gene cluster *cpk* (Takano, 2000; Takano, 2005). The methylenomycin Furans (MMFs) (**13**) (Figure 1.1) are signalling molecules (Corre *et al*., 2008), that specifically induce the production of methylenomycin in *S. coelicolor* A3(2). The MMFs are derived from the same metabolic precursors of GBLs and a key step in their biosynthesis is directed by a homologous enzyme.

The biosynthesis of MMFs is directed by an operon of three genes (*mmfLHP*) within the *mmy* gene cluster which is responsible of the biosynthesis of methylenomycin.

1.2.7 Germicidins

Germicidins (**14**) (Figure 1.1) have property of inhibiting spore germination and hyphal elongation in *Streptomyces* (Aoki *et al*., 2011) .The biosynthesis of five of them which involves a type III PKS were discovered in *S. coelicolor*. Germicidins catalyse the condensation reaction between specific β-ketoacyl-ACP thioester intermediates in fatty acid biosynthesis with ethyl- or methylmalonyl-CoA and then cyclised by a type III polyketide (Song *et al*., 2006). Similar to 2-Alkyl-4-hydroxylmythylfuran-3-carboxylic acid (AHFCAs) and GBLs, germicidins are derived from intermediates in fatty acid biosynthesis.

1.3 Fatty Acid Synthases (FASs)

In bacteria, fatty acid synthases (FASs) are multienzyme complexes that consist of individual highly conserved enzymes or domains (Black and DiRusso, 1994). Biosynthesis of fatty acids starts with acetyl-CoA resulting in fatty acids containing an even number of carbon atoms. The first step of fatty acid biosynthesis is the conversion of acetyl-CoA to malonyl-CoA in a reaction catalysed by acetyl CoA carboxylase (Alberts and Vagelos, 1972; Guchhait *et al*., 1974; Marini *et al*., 1995). The malonyl group is then loaded onto an acyl carrier protein (ACP) by malonyl-CoA: ACP transacylase (MCAT). An acyl-CoA unit (Acetyl-CoA in *E. coli* and isobutyl-, 2-methylbutyryl-, or isovaleryl-CoA in *Streptomyces*) is subsequently condensed with malonyl-ACP by synthase III (KASIII) to yield β-ketobutyryl-ACP. In some systems, the β-keto group is reduced by β-ketobutyryl-ACP reductase in a NADPH-dependent

reaction and then dehydrated by β-hydroxyl-ACP dehydratase. In the final step, enoyl-ACP reductase catalyses reduction to yield butyryl-ACP (McMurry and Begley, 2005).

KASII then condenses butyryl-ACP with another malonyl-ACP to give a β-keto-hexanoyl-ACP intermediate that undergoes reduction and dehydration reactions as before.

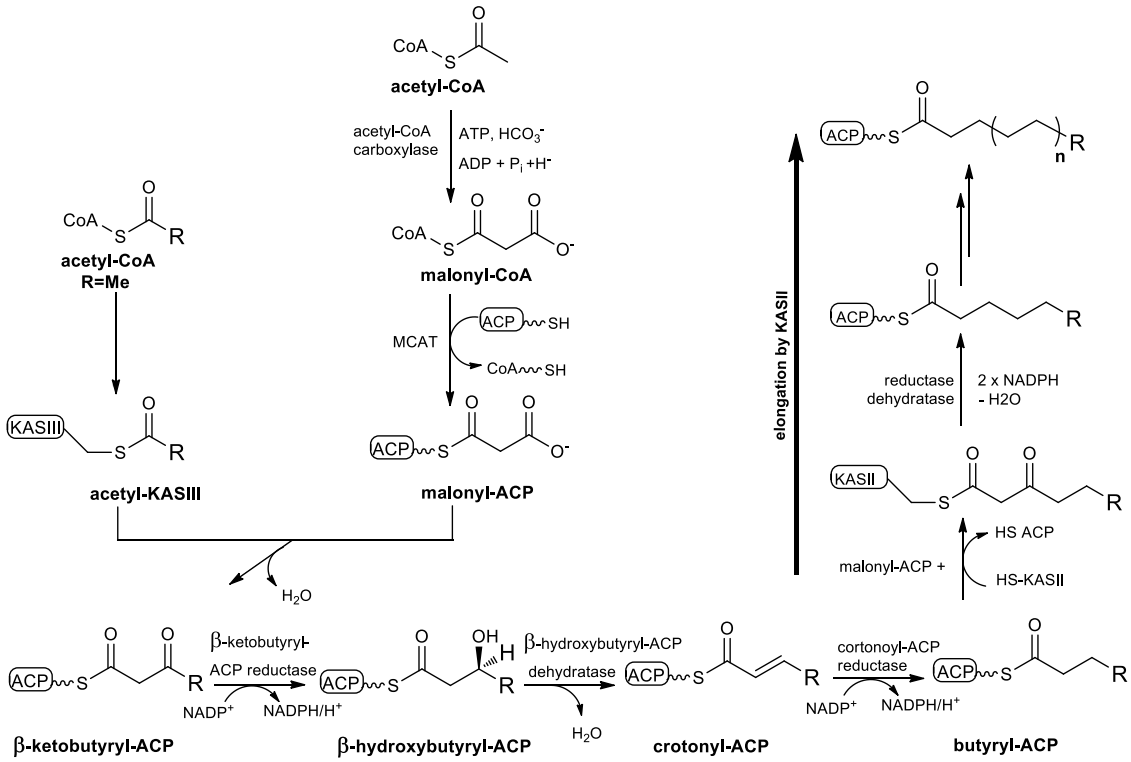


Figure 1.2 The biosynthesis of fatty acids in bacteria. In *E. coli* R= methyl, in *Streptomyces* R= isobutyryl-, 2-methylbutyryl-, isovaleryl-CoA.

The chain is extended each time with two more carbons until it reaches the required length by the action of KASII (Figure 1.2) (McMurry and Begley, 2005).

1.4 Polyketide Synthases (PKSs)

Polyketide synthases (PKSs) are large multienzyme systems responsible for catalysing the assembly of extraordinarily complex natural products from simple precursors (acetyl-CoA, propionyl-CoA, butyryl-CoA and their activated derivatives, malonyl-CoA, methylmalonyl-CoA or ethylmalonyl-CoA). Similar to fatty acids, polyketides are derived from the decarboxylative condensation of extender units. Carbon chain assembly and release are followed by further post-translational modifications, such as oxidation, methylation and glycosylation to give the final product (Figure 1.3) (McMurry and Begley, 2005). Differences in starter units selectivity, the number and selectivity of extender modules and the presence of additional domains, such as dehydratase, ketoreductase and acyl reductase domains within each extender module result in architecture differences between naturally occurring polyketides (Cortes *et al* 1995: McMurry and Begley, 2005).

PKSs are classified into three types. Type I PKSs are multifunctional proteins with individual functional domains while type II PKSs consists of complexes of individual mono-functional proteins. Type I can be further subdivided into iterative in which domains are reused and modular in which each domain is only used once. Type III PKSs directly use acyl-CoA substrates for chain extension with no involvement of an acyl carrier protein (ACP). There are some hybrid systems of type I/type II PKSs as well as hybrid type I PKSs/nonribosomal peptide synthetase known (Saxena *et al*., 2003: Foerstner *et al*., 2008).

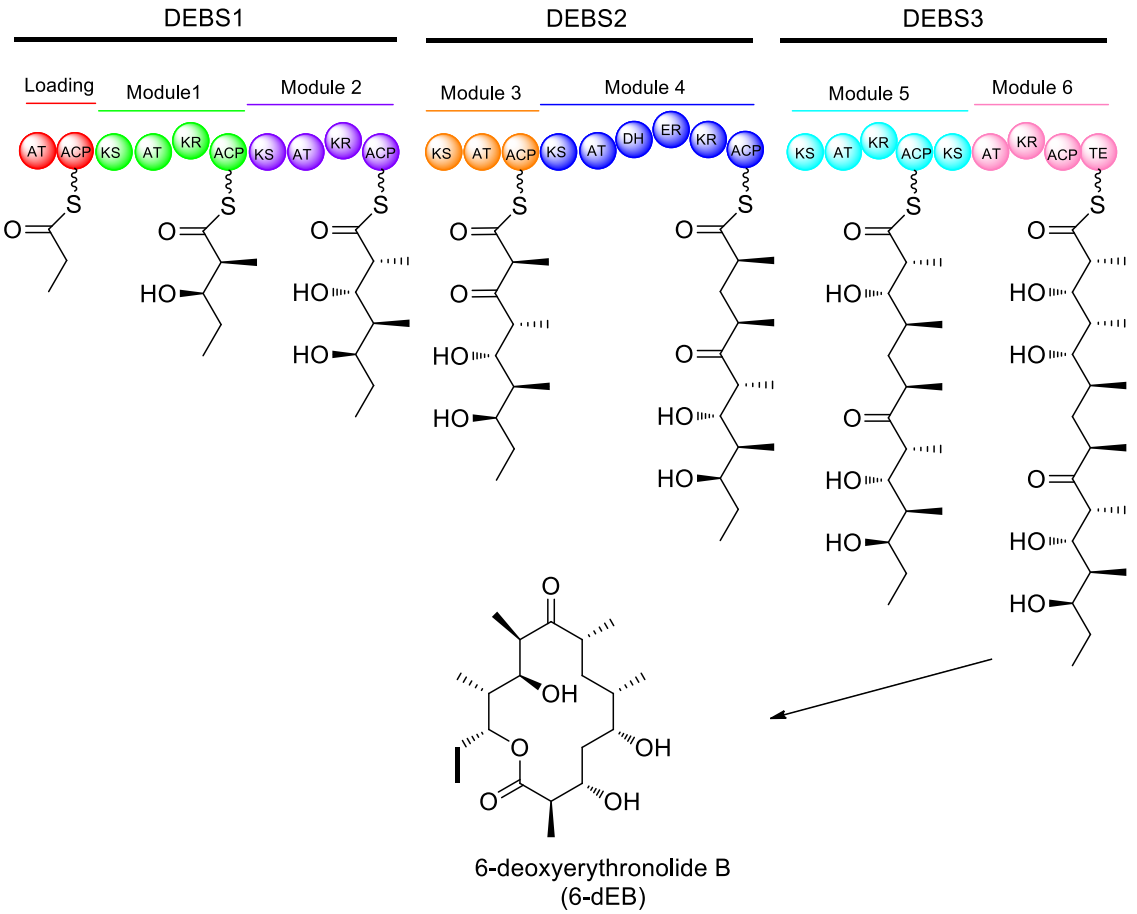


Figure 1.3 Proposed biosynthesis of erythromycin precursor 6-dEB and domain organization of the DEBS modular PKS. Domains are as follows: AT – acyltransferase, ACP – Acyl Carrier Protein, KS – ketosynthase, KR – ketoreductase, DH – dehydratase, ER – enoylreductase, TE – thioesterase.

The biosynthesis of erythromycin in *Saccharopolyspora erythraea* is the model example of a modular type I polyketide synthase (McGuire *et al*., 1952; Weber *et al*., 1990). Three modular type I PKSs, designated DEBS I, II and III are involved in the biosynthesis of 6 deoxyerythronolide (6 dEB), the polyketide core of erythromycin. They consist of a loading module to incorporate the first acyl group, six chain extension modules to add six further acyl groups to the polyketide chain, and a thioesterase domain to release the polyketide product. The loading module consists of an acyl transfer (AT) and acyl carrier protein (ACP) whilst a minimal module contains three

domains; an acyltransferase (AT), an acyl carrier protein (ACP) and a ketosynthase (KS). Propionyl-CoA (the acyl-CoA starter unit for erythromycin biosynthesis) is loaded by the AT domain of the loading module onto the adjacent ACP domain. Methylmalonyl-CoA (the extender unit for erythromycin biosynthesis) is loaded onto the ACP domain of module one by the AT domain within the same module and a Claisen condensation reaction is catalysed by the KS domain between this acyl-ACP and the acyl-ACP in the upstream loading molecule to extend the chain. This process is continued along the modules to give the growing chain. There are some additional domains in some extension modules like a ketoreductase (KR), which reduce the ketone to an alcohol, a dehydratase (DH) which dehydrates the resultant alcohol and an enoyl reductase (ER) which reduce the resultant double bond.

Finally, a thioesterase (TE) catalyses a cyclisation reaction to give the macrocyclic lactone 6 dEB (Figure 1.3) (Staunton and Wilkinson, 1997; McMurry and Begly, 2005) and releases the product from the multienzyme. This core structure is then modified by post PKS tailoring enzymes to give erythromycin.

1.5 Nonribosomal Peptide Synthetases (NRPSs)

Nonribosomal Peptide Synthetases (NRPSs) are multifunctional enzymes responsible for the biosynthesis of nonribosomal peptide specialised metabolites, produced mainly by bacteria and fungi (Challis, *et al*., 2000). Peptide-derived specialised metabolites can be biosynthesised via the condensation of proteinogenic and non-proteinogenic amino acids as well as D-amino acids to give peptides. NRPSs can, in many cases, work with PKSs to biosynthesise hybrid peptide-polyketide natural products (Ansari *et al*., 2004). A minimal NRPSs module consist of three core catalytic domains; adenylation (A) domain, thiolation (T) domain (PCP domain) and condensation (C) domain.

The loading module consist of A and PCP domains and the C-terminus of the final NRPSs module typically contains a thioesterase (TE) domain which catalyses the release of the peptide from the NRPS. Diversity of peptide products and modification to the growing peptide chain can be carried out by further specialised domains such as epimerisation (E), methylation (M) and reduction (R) domains (Finking and Marahiel, 2004). During nonribosomal peptide biosynthesis, the A domain selects the amino acid substrates, and activate it as an amino acyl adenylate, then transfers the activated amino acid to the peptidyl carrier protein (PCP) domain. The peptide bond formation between amino acids of two adjacent PCP domains is catalysed by the C domain (Garwin *et al*., 1980; Gehring *et al*., 1997). The TE domain of the last NRPS module catalyses either the hydrolysis or the cyclisation of the peptide to release it from the NRPS. (Donadio *et al*., 2007).

The biosynthesis of the siderophore coelichelin which was isolated from *S. coelicolor* (Challis and Ravel, 2000; Lautru *et al*., 2005) is an example of a natural product assembled by an NRPS. *cch* cluster is shown to encode for the required proteins for the biosynthesis of coelichelin. It was predicted that modules, 1, 2 and 3 of CchH incorporate L- δ-N-formyl-8δ-N-hydroxyornithine, L-threonine and L- δ-N-hydroxyornithine respectively.

The A, PCP and E domains of module 1 and the C domain of either module 2 or module 3 of CchH are used iteratively and other domains are skipped in the second iteration to form the tetrapeptide coelichelin (Figure 1.4).

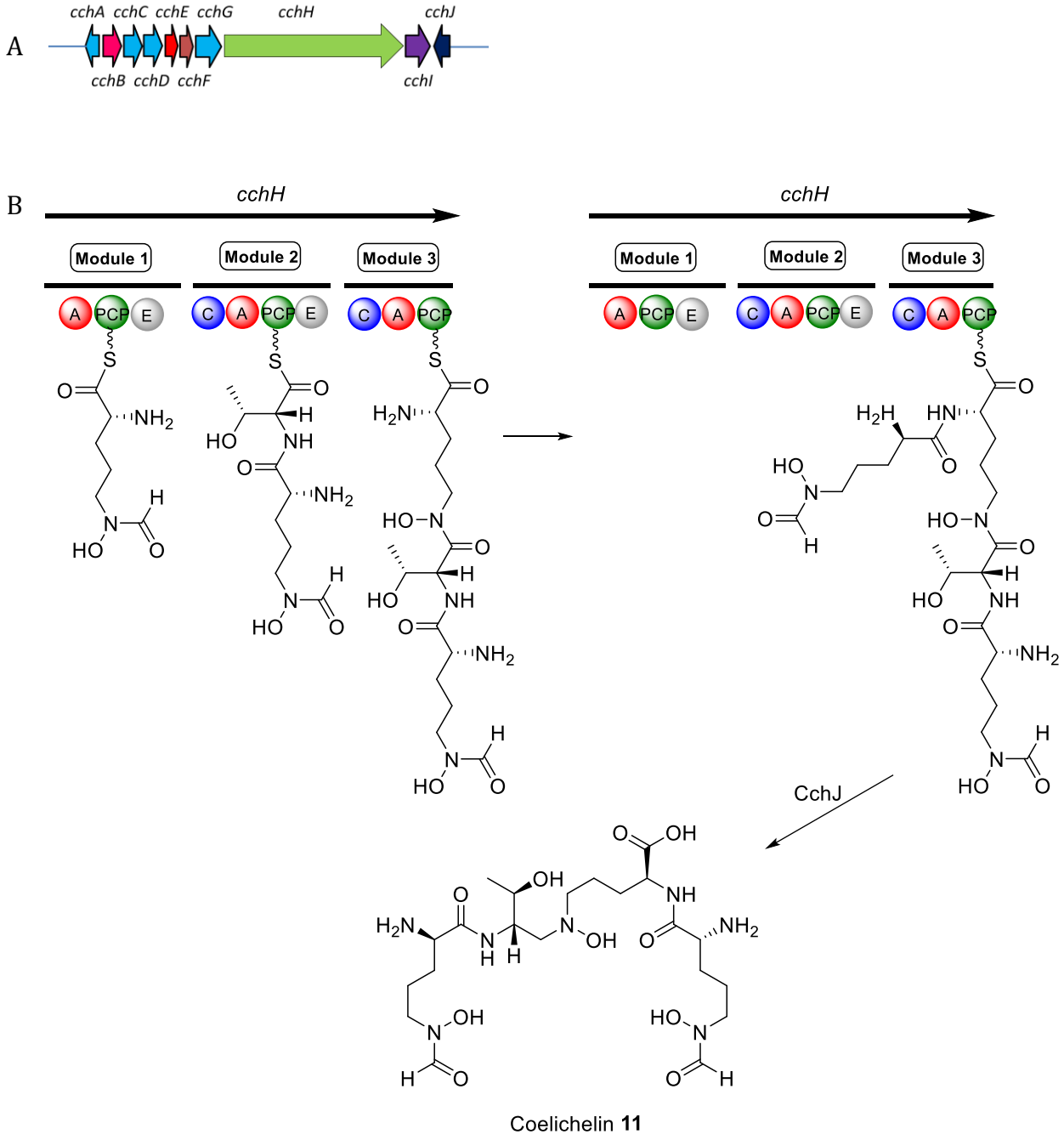


Figure 1.4 A. Organisation of *cch* gene cluster. B. The module and domain organization of the NRPS encoded by *cchH*. Domains: C – condensation, A - adenylation, PCP- thiolation, and E- epimerisation.

The product is released from the NRPS by CchJ catalysed hydrolysis. (Lautru *et al.*, 2005).

1.6 Prodiginines

Prodiginines are a family of red-pigmented pyrrole-containing specialised metabolites produced by actinomycetes and eubacteria. The colour produced by these bacteria resembles droplets of blood and has been linked to many historical events such as in the "bleeding" of bread during Alexander the Great's siege of Tyre in 332 B.C. This was interpreted as a good omen for victory and was used to justify the bloodshed of inhabitants (Furstner, 2003). Now it is known that this "blood" is caused by living organisms, mainly *Serratia marcescens*, from which the first prodiginine, prodigiosin, was isolated (Figure 1.5) (Gaughran, 1969; Bennett and Bentley, 2000).

The family of prodiginines contain a common highly conjugated tripyrrole system in which the rings are designated A, B and C (Figure 1.6). This conjugated tripyrrole system is responsible for their characteristic intense red colour. Prodigiosin (**18**), the first member of the prodiginine family to be discovered, was first isolated in 1902 from *Serretia marcescens*, but not structurally elucidated until 1960 (Bentley *et al*., 2002).

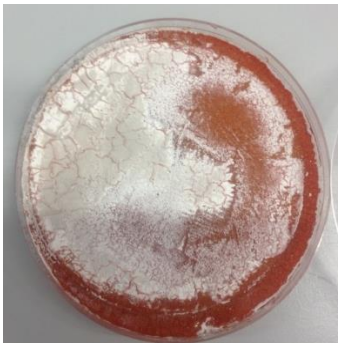


Figure 1.5 Plate of *S. longispororuber* producing red-pigmented prodiginine.

Since that discovery, a large number of prodiginines produced by actinomycetes have been isolated including undecylprodigiosin (**2**) produced by *S. coelicolor* and *S. longispororuber* (Wasserman *et al*., 1969; Gerber, 1975), and its carbocyclic derivatives (with unusual *ansa*-bridged rings incorporating on the pyrrole C ring) streptorubin B (**3**) (produced by *S. coelicolor*) (Mo *et al*., 2008) and metacycloprodigiosin (**19**) (produced by *S. longispororuber*) (Wasserman *et al*., 1969).

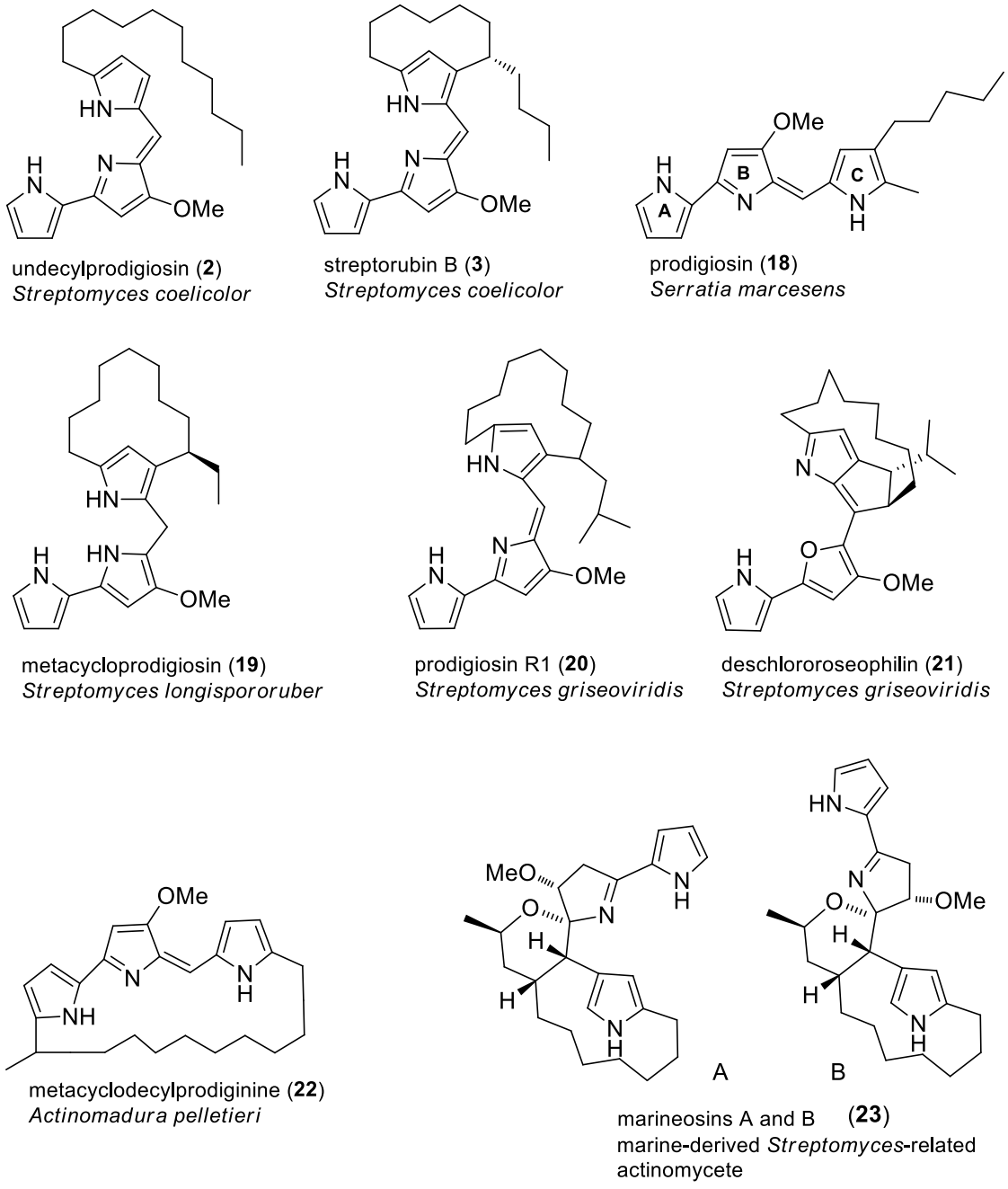


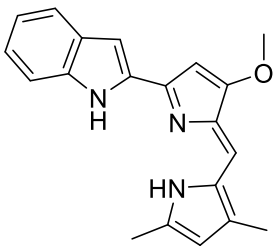
Figure 1.6 Structures of prodiginines natural products.

Among the prodiginines there are several isomeric cyclic derivatives such as metacyclodecylprodiginine (**22**) (Gerber, 1975) produced by *Actinomadura pelletieri* in which an alkyl chain is attached to the A and C rings forming part of a macrocycle. Another member of the prodiginine family is prodigiosin R1 (**20**) which was isolated

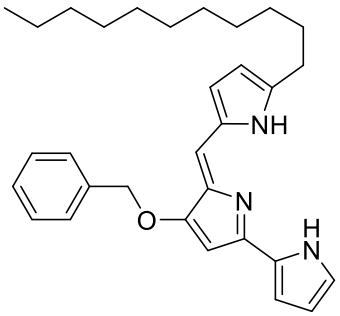
from *Streptomyces griseoviridis*, together with roseophilin (**21**) which contains a methoxyfuran rather than a methoxypyrrole B-ring and a chloro substituent on the A ring. Marineosins (**23**) are other structurally related products from a marine actinomycete (Boonlarppradab *et al* 2008) (Figure 1.6)

Prodiginines have received attention due to their broad range of biological activities. In addition to their antibacterial activity, prodiginines also have potent immunosuppressive, antimalarial and anticancer properties (Williamson *et al*., 2007).

Many chemically synthesised analogues of undecylprodigiosin (**2**) have been shown to possess immunosuppressant activity at non-toxic doses (Tsuji *et al*., 1992; Magae *et al*., 1996; Lee *et al*., 2000). **2** can inhibit the proliferation of human T-cells and suppress T-cell-dependent antibody responses without damaging the lymphoid organs (Azuma *et al*., 2000). Additionally, PNU-156804**,** an analogue of (**2**) (Figure 1.7) was shown to be equally effective but less toxic than the natural product (Mortellaro *et al*., 1999).



PNU-156804



Obatoclax GX15-070

Figure 1.7 Structures of PNU-156804 (a synthetic prodiginine analogues) and Obatoclax.

The potential anticancer activity of the prodiginines has also been studied. They showed anticancer activity against many cell lines, including lung, colon, kidney and breast with little cytotoxicity against noncancerous cells. Prodigiosin (**18**) produced by *Serratia marcescens* was used in preclinical trials for the treatment of pancreatic cancer (Zhang *et al*., 2005). Furthermore, a synthetic analogue of streptorubin B and metacycloprodigiosin, obatoclax (GX15-070), (Figure 1.7) has undergone phase II oncology trials for the treatment of chronic leukemia, lymphoma, lung cancer and other solid tumors (Trudel *et al*., 2007). Obatoclax interact with anti-apoptotic members of the Bcl-2 family of proteins, which have been shown to be overexpressed in numerous cancer cell lines, preventing normal cell death from occurring. Obatoclax binds to the conserved binding site of these anti-apoptotic proteins, resulting in the release of the pro-apoptotic proteins, restoring the normal cell death process (Danial and Korsmeyer, 2004; Nguyen *et al*., 2007; Nguyen *et al*., 2015).

In addition to their anticancer properties, prodiginines have anti-malarial activity. Screening against *Plasmodium falciparum* the causal agent of malaria, showed that **2**, **19** and **3** are more potent than chloroquine (Papireddy *et al*., 2011), although the mode of action is unknown (Montaner *et al*., 2000). Moreover, several synthetic analogues of these prodiginines demonstrated even greater potency and were even shown to cure induced malaria disease in several mice studies (Papireddy *et al*., 2011). Very recently, obatoclax showed prevention of viral fusion by inhibition of endosomal acidification (Varghese *et al*., 2017).

1.7 Early studies into prodiginine biosynthesis

Early studies into the biosynthesis of prodiginine were carried out on prodigiosin (**18**) in *Serratia marcescens* with later studies focusing on undecylprodigiosin (**2**) and metacycloprodigiosin (**19**) in *S. longispororuber* (Wasserman 1969). Feeding experiments using ^{13}C labelled precursors and isolation of putative biosynthetic intermediates led to the hypothesis that in *S. coelicolor* **2** is formed via the condensation of 2-undecylpyrrole (2-UP) (**15**) and 4-methoxy-2,2' bipyrrole-5-carboxaldehyde (MBC) (**16**) (Cerdeño *et al*., 2001). Structural similarities between **18** and **2** suggested that they would share **16**, as a common intermediate in their pathway. Condensation of **16** with either 2-methyl-3-pentylpyrrole (MAP) (**17**) or **15** would give prodigiosin **18** and **2**, respectively (Figure 1.8) (Morrison, 1966; Wasserman, 1973).

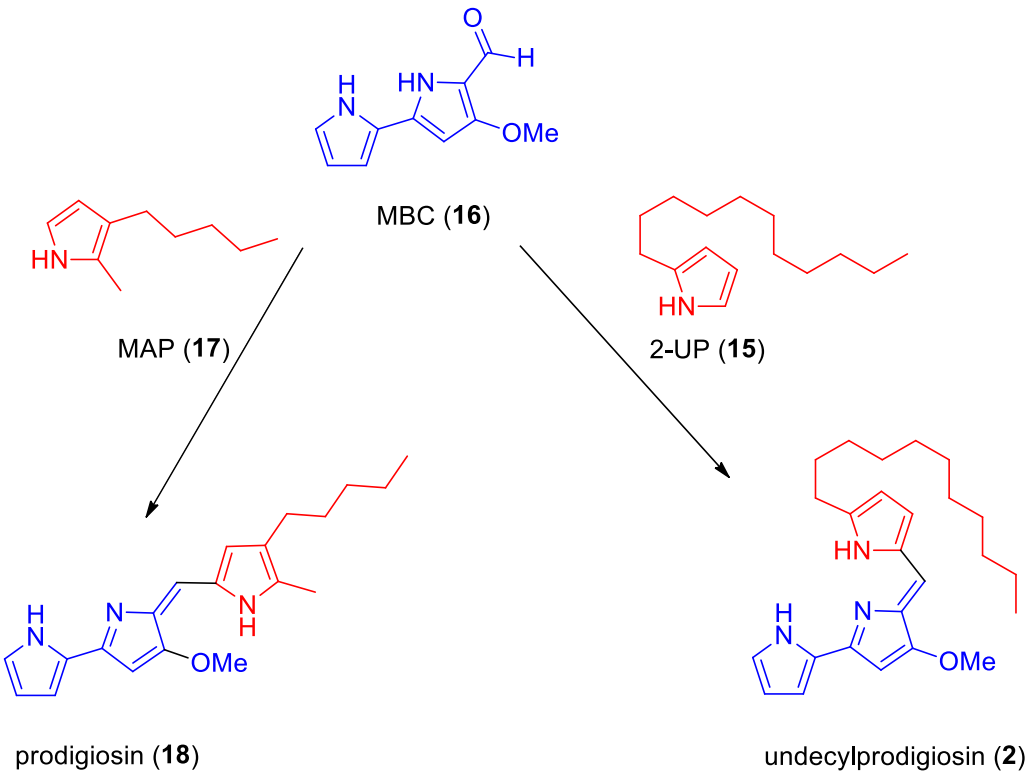


Figure 1.8 MBC as a common intermediate in prodigiosin (18) and undecylprodigiosin (2) biosynthesis.

It was proposed that the biosynthesis of undecylprodigiosin and prodigiosin is a bifurcated process involving the condensation of **16** with a mono pyrrole **15** or **17** respectively (Cerdeño *et al*., 2001; Stanley *et al*., 2006; Williamson *et al*., 2005).

The feeding of isotopically labelled precursors demonstrated that **16**, in both systems, is generated from L-proline, L-serine, L-methionine and acetate (Gerber *et al*., 1978; Wasserman et., al 1973; Wasserman et., al 1974). It was proposed that **19**, in *S. longispororuber* is formed from **2** via an oxidative carbocyclisation reaction. As the only difference between streptorubin B (**3**) and metacycloprodigiosin (**19**) is the size of their macrocycles, it was proposed by Challis *et al* that **3** (Butyl-metacycloheptylprodigiosin) was formed in an analogous manner.

1.8 Streptorubin B biosynthetic gene cluster

The biosynthetic gene cluster of a prodiginine produced by *S. coelicolor* (the *red* cluster) was first identified in the 1970s (Rud and Hopwood, 1980), then cloned and expressed in a heterologous host in 1990 (Malpartida *et al*., 1990). In 2002, the entire genome sequence of *S. coelicolor* A3(2) was published (Bentley *et al*., 2002) and the gene cluster responsible for the biosynthesis of undecylprodigiosin (**2**) and streptorubin B (**3**) was identified. The *red* cluster (Figure 1.9) contains twenty three genes, arranged in four transcription units. The function of the encoded proteins of most genes was proposed on the bases of sequence comparison with proteins of known functions (Cerdeño *et al*., 2001).

The role of these enzymes has subsequently been revised in the light of subsequent experimental evidence mainly from deletion experiments (Stanley *et al*., 2006; Haynes *et al*., 2008; Mo *et al*., 2008).

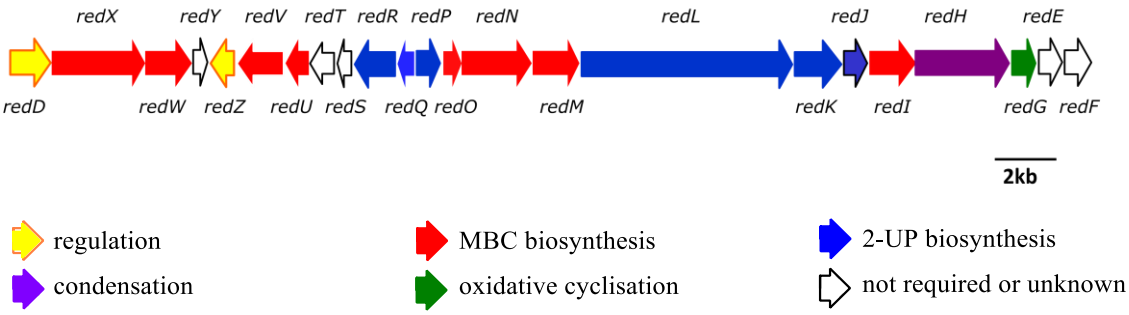


Figure 1.9 red gene cluster in *Streptomyces coelicolor* and the function of the encoded genes.

redD and *redZ* are known to encode pathway specific regulators. Eight genes have been shown to be required for the biosynthesis of the key intermediate **16** (*redI, redM, redN, redO, redU, redW, redX and redV*) (Williamson *et al*., 2005; Stanley *et al*., 2006) and six genes are required for the biosynthesis of **15** (*redJ, redL, redP, redR, redQ and redK*). *redH* encodes the enzyme that catalyses the condensation of the two late intermediates **15** and **16** (Haynes *et al*., 2008). *redG* encodes a Rieske oxygenase that catalyse the oxidative carbocyclisation of **2** to form **3** (Stanley *et al*., 2006) . Three genes (*redS*, *redF*, and *redE*) do not appear to be required for the biosynthesis of prodiginines.

Within the cluster responsible for prodiginine biosynthesis, there are two genes that encode PKS-like enzymes involved in MBC biosynthesis; RedX and RedN. RedX contains two ketosynthase domains which are similar to each other (29%) and to the KS domain of type I modular PKSs (~30%). The N-terminal RedX KS domain contains aspartate in the conserved active site which the C-terminal KS domain has a cysteine residue in this position. RedN consist of two ACP domains with serine residues in the N-terminal of RedN and a C-terminal α-oxoamine synthase (OAS) domain (Cerdeño *et al*., 2001; Stanley *et al*., 2006).

One PKS-NRPS hybrid enzyme (RedL) is involved in 2-UP biosynthesis. RedL consist of six domains. The N-terminal A domain is homologous to an NRPS adenylation domain. The A domain is followed by an ACP domain, a KS domain, an AT and another ACP domain, all of which are homologous to the corresponding domains in type I PKSs. The C-terminal α-oxoamine synthase (OAS) domain of RedL showed 34% identity to the C-terminal (OAS) domain of RedN (Cerdeño *et al*., 2001; Mo *et al*., 2008).

1.9. Biosynthesis of streptorubin B

1.9.1 2-UP biosynthesis

Previous studies on prodiginine biosynthesis suggested that 2-UP was biosynthesised by condensation of one unit of β-ketomyristoyl thioester produced from seven units of acetate and one unit of glycine (Wasserman *et al*., 1973; Gerber, *et al*., 1978).

In *S. coelicolor*, FabH (KASII), FabF (KASII) and FabC (ACP) analogues (components of fatty acid synthase -FAS) encoded by three genes from the *red* cluster, *redP*, *redR* and *redQ* respectively, were proposed to initiate 2-UP biosynthesis by generating dodecanoic acid (Cerdeño *et al*., 2001).

Deletion of *redP*, *redR* and *redQ* (Mo *et al*., 2005; Mo *et al*., 2008) leads to the reduction but not abolition of the production of prodiginine suggesting that FabH and FabF (Fatty acid biosynthetic enzymes) could catalyse the elongation steps of dodecanoyl-RedQ biosynthesis but efficient and selective prodiginine biosynthesis requires *redP*, *redR* and *redQ*. This deletion also led to production of branched chain prodiginine analogues derived from isovaleryl, isobutyryl and 2-methylbutyryl starter

units (Mo *et al*, 2005 and Mo *et al*, 2008), presumably because the FAS enzymes potentially make these precursor fatty acids.

RedP is proposed to catalyse the condensation of an acetyl-CoA starter unit with a malonyl-RedQ (ACP). Type II FAS enzymes that reduce the resulting acetyl thioester to butyryl-ACP by ketoreduction, dehydration and enoyl-reduction. Subsequent elongation steps with malonyl-RedQ catalysed by RedR, and reductions catalysed by FAS enzymes result in the formation of dodecanoyl-RedQ (Figure 1.10) (Cerdeño *et al*., 2001; Mo *et al*., 2005; Mo *et al*., 2008).

It has been shown that transfer of the dodecanoyl group to the first ACP domain of RedL is catalysed by RedJ via hydrolysis to give dodecanoic acid (Whicher *et al*., 2011), which is subsequently loaded onto the first RedL ACP domain after activation as an adenylate by the adenylation (A) domain of RedL. The dodecanoyl thioester on the first ACP domain of RedL is condensed with a malonyl group attached to the second ACP, in a reaction catalysed by the KS domain, giving β-ketomyristoyl-ACP. The pyridoxal 5-phosphate (PLP)-dependent OAS domain of RedL then catalyses the condensation of glycine with the attached intermediate.

Release of the intermediate and cyclisation provides 4-keto-2-undecylpyrroline (Figure 1.10) (Whicher *et al*., 2011). In the final step, the reduction and dehydration of the keto group is catalysed by NAD(P)H dependant enzyme RedK to give **15**.

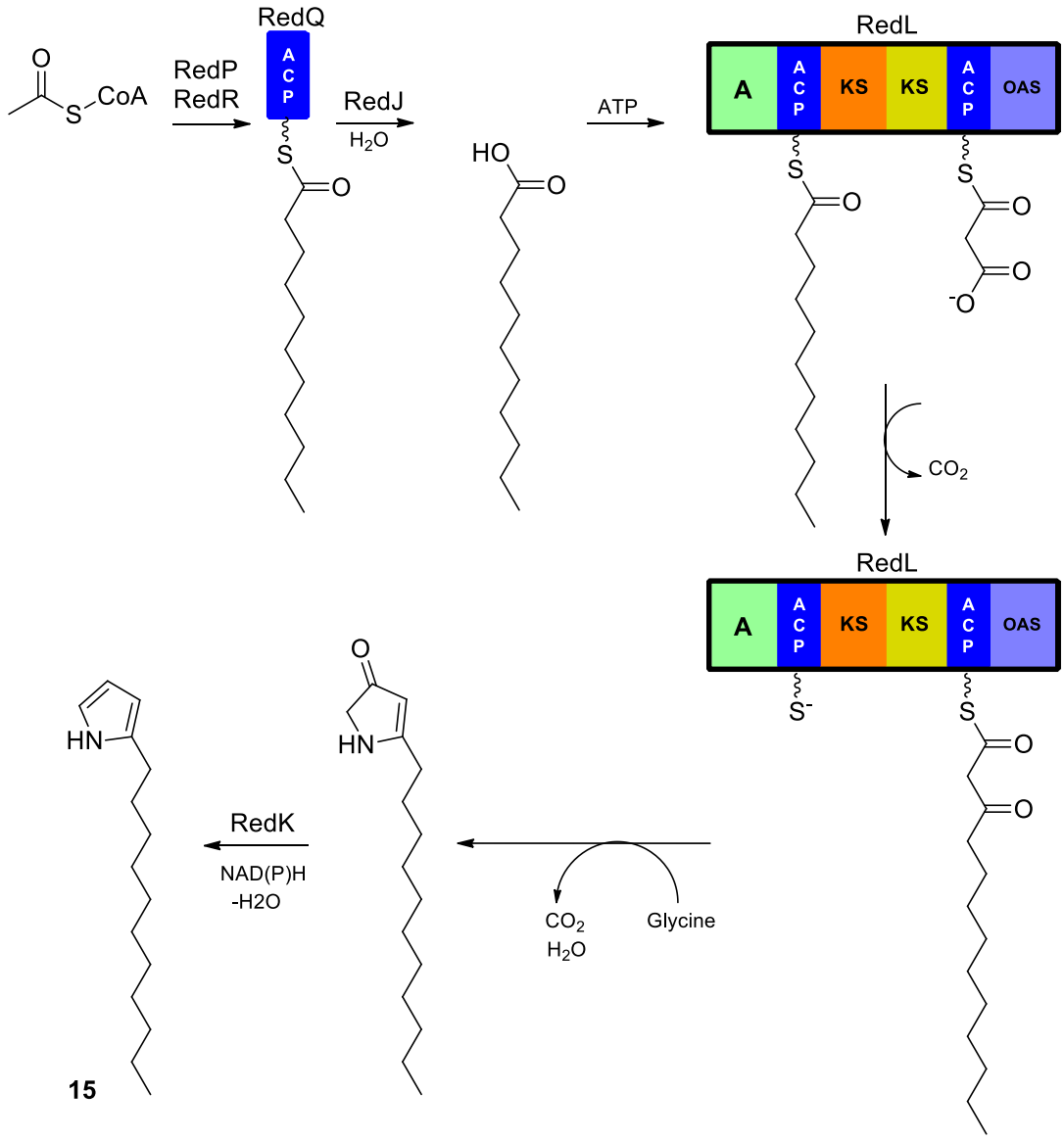


Figure 1.10 The biosynthetic pathway to 2-undecylpyrrole (**15**).

1.9.2 MBC biosynthesis

The mechanism for biosynthesis of **16** in *S. coelicolor* was initially proposed by Cerdeño *et al* (2001) and Thomas *et al* (2002), and then a similar mechanism was proposed for biosynthesis of **16** in *Serratia* sp. by Harris *et al* (2004) and was revised by Williamson (2006). The biosynthesis of **16** starts with post-translational

phosphopantetheinylation of the peptidyl carrier protein (PCP) RedO catalysed by RedU, a phosphopantetheinyl transferase (PPTase) (Figure 1.11). RedM catalyses the conversion of the carboxyl group of L-Proline to form an amino-acyl adenylate, which is then loaded onto the phosphopantetheine arm of the RedO- PCP. Pyrrolidine is dehydrogenated by RedW in a FAD-dependent reaction and a pyrrole ring is generated (Thomas, 2002). The resulting pyrrole 2-carboxyl intermediate is then transferred to the C-terminal ketosynthase (KS) domain of RedX. Pyrrole-2 carboxyl-RedX is then condensed with a malonyl unit attached to one of the ACP domain of RedN giving a β-keto-ACP- thioester. This is followed by decarboxylative condensation of L-Serine with this β-keto-ACP- thioester catalysed by the pyridoxal-5-phosphate (PLP) dependent α-Oxoamine synthase (OAS) domain of RedN resulting in product release from the ACP. Cyclisation and dehydration of the resulting amino diketone provides 4-hydroxy-2, 2'-bipyrrole-5 methanol (HBM) (Figure 1.11).

The biosynthesis of **16** is completed by oxidation of the primary alcohol to the corresponding aldehyde, catalysed by RedV (unpublished data from Challis group) and SAM-dependent methylation of the pyrrole hydroxyl group, catalysed by RedI.

Strains lacking RedT and RedY produce significantly less prodiginines which can be restored by feeding chemically synthesised **16** to the mutant strains, but their function is currently unknown.

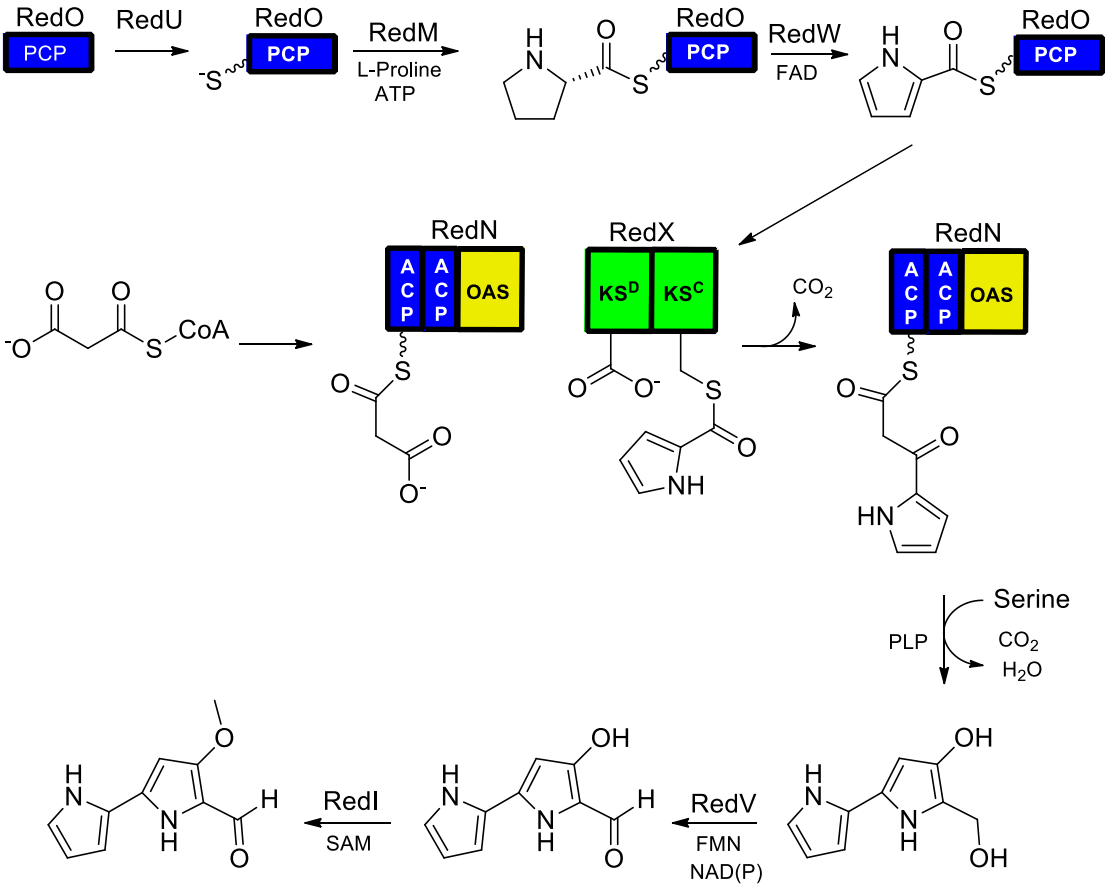


Figure 1.11 The biosynthetic pathway to methoxy-2, 2' bipyrrole-5-carboxaldehyde (MBC) (**16**).

1.9.3 Condensation of 2-UP and MBC

Originally, it was proposed that the condensation of 2-UP (**15**) and MBC (**16**) could be spontaneous *in vivo* and no enzyme was required to catalyse this reaction as it can happen in the presence of a Bronsted acid (Wasserman *et al*., 1969; Cerdeño *et al*., 2001). However, investigation of a *S. coelicolor* mutant lacking RedH produces no undecylprodigiosin or streptorubin B but accumulation of **15** and MBC indicating that RedH catalyses the condensation of **15** and **16** (Haynes *et al*., 2008). Additionally, investigation of RedH analogue PigC in *Serratia marcescens* showed that it is required for the condensation of **16** and MAP to give prodigiosin (Williamson *et al*., 2005). A

heterologous host expressing RedH in *Streptomyces venezuelae* fed with chemically synthesised **15** and **16** results in the production of undecylprodigiosin but not streptorubin B, showing that RedH is only required for catalysing the condensation reaction of **15** and **16** (Figure 1.12) (Sydor *et al*., 2011).

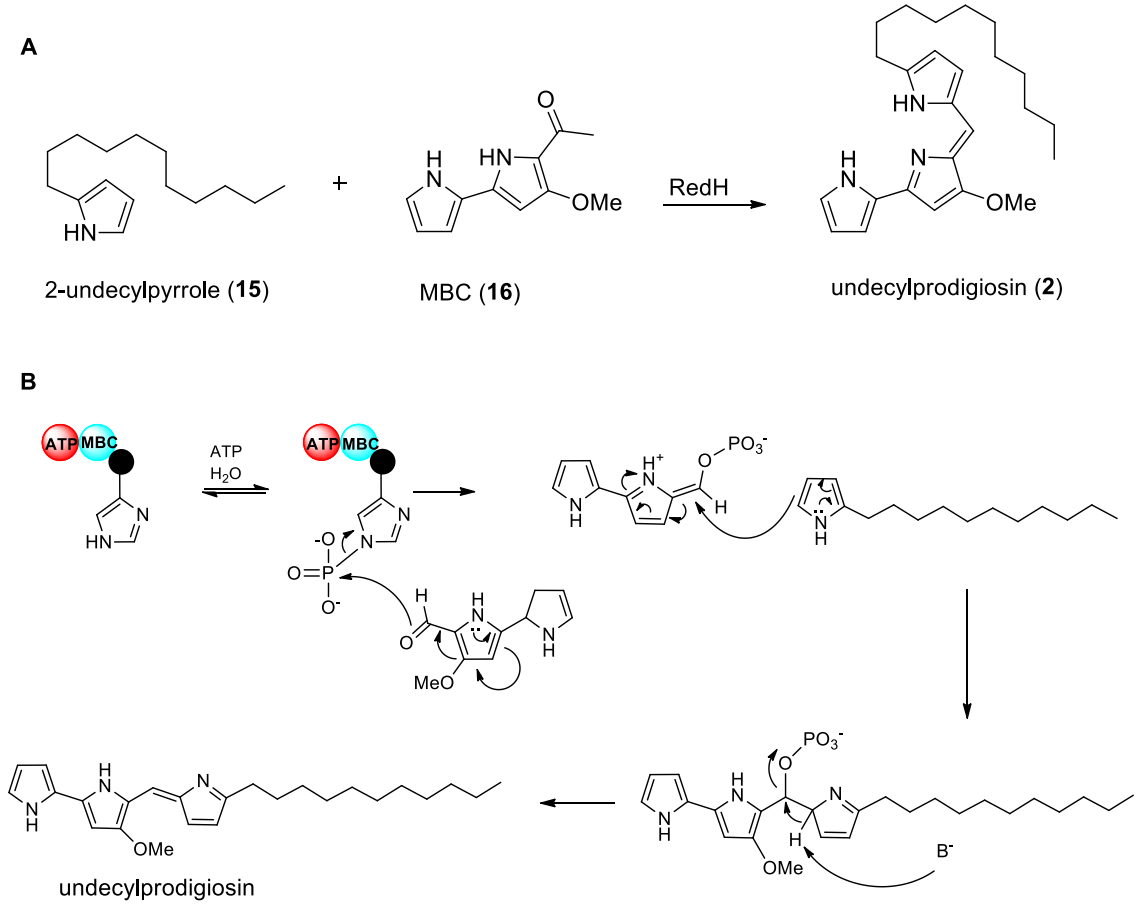


Figure 1.12 A. Condensation of 2-UP (**15**) and (MBC) (**16**) to form undecylprodigiosin (**2**) catalysed by RedH in *S. coelicolor*. B. proposed condensation mechanism catalysed by RedH.

Comparison of RedH with proteins of known function suggest it is a phosphotransferase with three functional domains; the C-terminal domain is homologous to the phosphotransferase domain of phosphate pyruvate dikinase (PPDK) and the N-terminal domain is homologous to the ATP binding phosphotransferase domain of phosphoenol

pyruvate synthase (PEPS) (Cerdeño *et al*., 2001 and Mo *et al*., 2008). The third domain of RedH does not show any sequence similarity to other proteins with known function and is proposed to bind **16**.

Based on the known mechanism of PPDK and PEPS it was proposed that RedH activates the carbonyl group of **16** by phosphorylation of the oxygen atom in an ATP-dependent reaction. Subsequent nucleophilic attack of **15**, followed by elimination of phosphate and rearomatisation gives undecylprodigiosin (Figure 1.11).

1.9.4 Oxidative carbocyclisation of undecylprodigiosin

The last step of the biosynthesis of prodiginine in *S. coelicolor* is the oxidative carbocyclisation of undecylprodigiosin (**2**) to form the 10-membered *ansa*-bridged carbocycle, streptorubin B (**3**) (Figure 1.6). It was hypothesised that the Rieske-non-haem iron-dependent RedG catalyse this reaction (Cerdeño, 2001). Oxidative cyclisation reactions are key steps in the biosynthetic pathway of many natural products e.g. penicillins, fosfomycin and clavulanic acid (Konomi *et al*., 1979; Elson *et al*., 1987; Hammerschmidt, 1991; Seto *et al*., 1991; Zerbe *et al*., 2004). The enzymes that catalyse the reaction usually contains non-haem-iron cofactors and use molecular oxygen as a co-substrate (Seto *et al*., 1991; Roach *et al*., 1995; Zhang *et al*., 2000; Liu *et al*., 2001; Higgins *et al*., 2005).

A mutant of *S. coelicolor* lacking *redG* produced **2** but not **3** implicating RedG in this carbon-carbon bond forming reaction. A heterologous host *S. venezuelae* expressing both *redH* and *redG*, fed with chemically synthesised 2-UP (**15**) and MBC (**16**) resulted in the production of undecylprodigiosin and streptorubin, indicating that RedG is responsible for catalysing the oxidative carbocyclisation reaction of **2** into **3** (Sydor *et*

al, 2011). Feeding *S. coelicolor* mutants lacking genes required for 2-UP (**15**) biosynthesis with synthetic **15**, restores the production of undecylprodigiosin and streptorubin B demonstrating that oxidative cyclisation must occur after assembly of **15** (Mo *et al.*, 2008). However, it is still unknown whether carbocyclisation in streptorubin biosynthesis occurs before or after condensation with MBC (**16**).

The sequence of RedG compared with proteins of known function showed a strong similarity to the Rieske non-haem-iron dependent-enzyme, naphthalene dioxygenase (NDO) (Figure 1.12). McpG, a RedG homologue from *S. longispororuber* also revealed a high similarity to RedG and NDO.

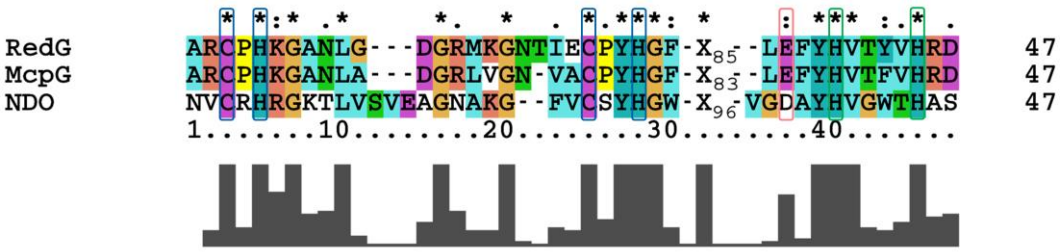


Figure 1.12: Amino-acid sequence comparison of RedG (*S. coelicolor*), McpG (*S. longispororuber*) and the Rieske non-haem iron dependent enzyme naphthalene dioxygenase (NDO); residues that ligate the [2Fe-2S] cluster and Fe(II) atom – are highlighted in blue and green respectively; a mutation from aspartate in NDO to glutamate in RedG and McpG is highlighted in red.

1.10 Rieske oxygenases.

Rieske non-haem iron dependent oxygenases catalyse a wide range of reactions such as biodegradation of xenobiotics and biosynthesis of bioactive natural products. Rieske oxygenases were first identified as enzymes responsible for degradation of many aromatic compounds by *Pseudomonas putida* (Axcell and Geary, 1975; Gibson *et al.*,

1968). These enzymes catalyse the first step in the oxidative degradation of aromatic compounds *via* regio- and stereospecific cis-dihydroxylation to generate dihydrodiols.

Naphthalene dioxygenase (NDO) is a well-characterised Rieske oxygenase enzyme which catalyse the first step in the degradation of naphthalene through *cis*-1,2-dihydroxylation (Figure 1.13) (Eaton and Chapman, 1992). The structure of NDO was elucidated by X-ray crystallography and consists of a three component system containing a catalytic α-subunit and a structural β-subunit organised into an $\alpha_3\beta_3$ complex (Figure 1.13) (Kauppi *et al* 1998; Ferraro *et al*., 2005).

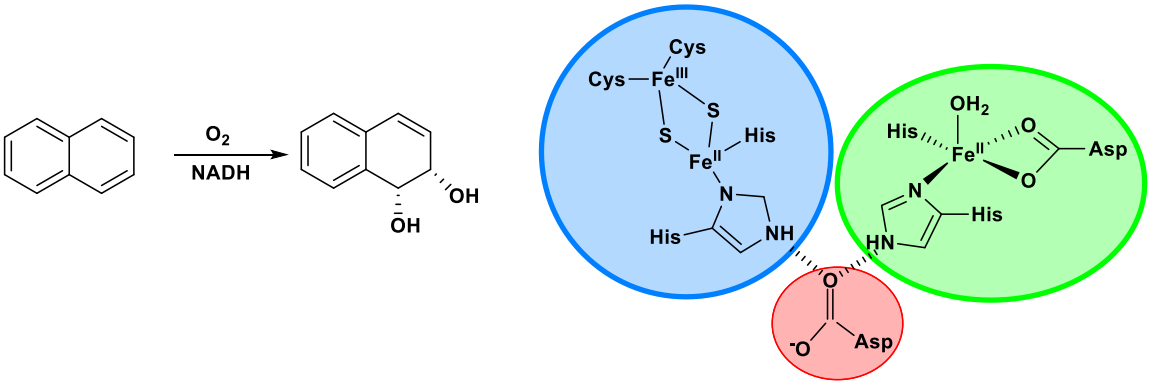


Figure 1.13 Reaction catalysed by naphthalene dioxygenase (left), proposed [2Fe-2S] Rieske cluster (blue), non-haem iron active site (green) and bridging aspartate residue (red).

The [2Fe-2S] cluster within the N-terminal Rieske domain of the α-subunit is coordinated to its protein by two histidine residues and two cysteine residues within the conserved sequence motif $CXHX_{17}CX_2H$ (Figure 1.13). The C-terminal domain contains an iron atom which is coordinated by two histidine residues and an aspartate residue (2-His-1-carboxylate-triad) which is characteristic for a large number of non-haem iron containing enzymes (Jiang *et al*., 1996).

NDO utilises molecular oxygen and NADH as cosubstrate (Figure 1.13). Using NIR-MCD spectroscopy, the non-haem iron centre of NDO has been characterised by Salomon and co-workers (Ohta *et al*., 2008). Electrons are transferred from NADH to the Rieske [2Fe-2S] cluster, typically *via* a flavin-dependent ferredoxin reductase and a separate ferredoxin. These electrons are transferred to the non-haem iron centre within the active site of the enzyme, where the oxygen is bound and the catalytic cycle takes place and are used for dioxygen reduction (Bugg and Ramaswamy, 2008; Ferraro *et al*., 2005). Several examples of Rieske oxygenases that catalyse a variety of catabolic and biosynthetic reactions were recently discovered.

KshAB is an example of Rieske oxygenases from *Mycobacterium tuberculosis* encoded by a gene cluster responsible for cholesterol catabolism, which may play a role in pathogenesis. (Van der Geize *et al*., 2007).

DAF is another cholesterol-metabolising Rieske oxygenase which has been reported by Niwa and co-workers (Van der Geize *et al*., 2007). The gene encoding this enzyme is conserved in nematodes and insects and its deletion is lethal.

NdmA and NdmB, two Rieske oxygenases from *P*. *Putida* CBBS have recently been characterised by Subramanian and co-workers. NdmA and NdmB are also involved in catabolic pathways by catalysing caffeine degradation through oxidation of N_1-demethylation and N_3-demethylation, respectively (Summers *et al*., 2012). NdmD, a NADH-dependent reductase, is able to supply electrons to the [2Fe-2S] centre of NdmA and NdmB.

PrnD is a Rieske oxygenase that is involved in the biosynthesis of pyrrolnitrin, a broad spectrum antifungal produced by several *Pseudomonas* and *Burkholderia* species. PrnD

which was biochemically characterised by Zhao and co-workers (2005) catalyses the oxidation of an aniline derivative to the corresponding nitrobenzene derivative.

1.11 RedG as a novel member of Rieske Oxygenases family

An amino acid sequence alignment with the naphthalene dioxygenase α-subunit reveals that the N-terminal domain of RedG contains the His and Cys residues that bind the [2Fe-2S] cluster as in the N-terminal domain of NDO (Kauppi *et al*., 1998) and the C-terminal domain contains the two His residues of the 2-His-1-carboxylate iron-binding triad in the non-haem centre. The aspartate residue implicated in electron transfer from the [2Fe-2S] centre to the non-haem iron centre in NDO is mutated to glutamate (Figure 1.13) (Sydor *et al*., 2011).

RedG, as a Rieske oxygenase enzyme requires a reductase and ferredoxin which here are not encoded by genes within the *red* cluster (Cerdeño, 2001). However, the six ferredoxin and three reductases encoded outside of the *red* cluster could mediate electron transfer from NAD(P)H to RedG enzyme (Lei *et al*., 2004).

1.12 Stereochemical and mechanistic studies of the RedG-catalysed reaction

Based on the catalytic mechanism of NDO, a mechanism for the RedG-catalysed oxidative carbocyclisation reaction has been proposed (Figure 1.14) (Barry and Challis, 2013). The proposed mechanism involves initial binding of the substrate to the active site resulting in loss of water from non-haem iron centre. Oxygen binding and transfer of an electron from the (2-Fe-2S) cluster to the non-haem centre results in formation of a peroxide complex.

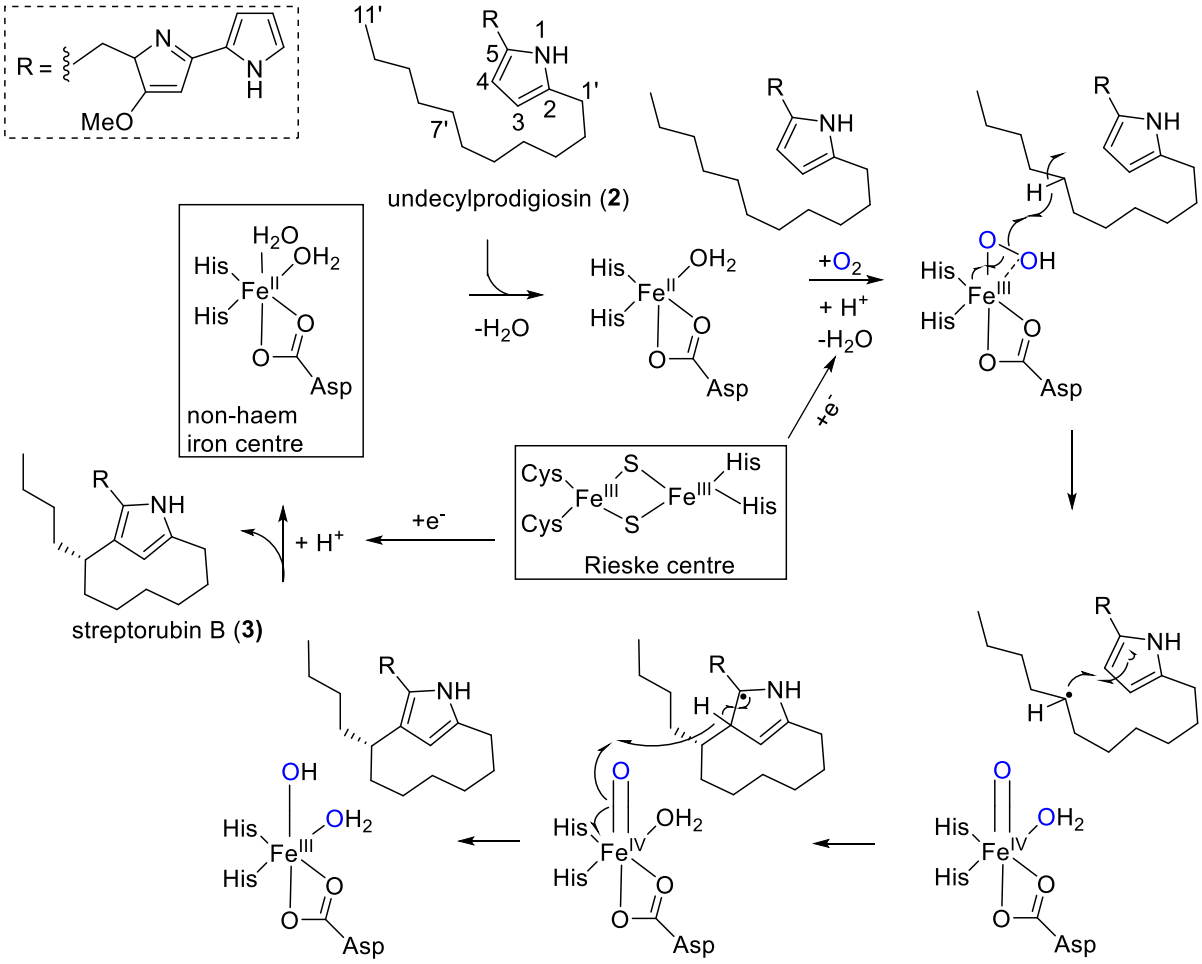


Figure 1.14 Proposed mechanism for the RedG-catalysed oxidative carbocyclisation of undecylprodigiosin (**2**) to streptorubin B (**3**).

The complex undergoes protonation and loss of water forming a bidentate ferric hydroperoxide complex. Cleavage of the O-O bond coupled with hydrogen atom abstraction from C-7 of the substrate forms a Fe(IV)=O(OH_2) complex and a secondary alkyl radical which adds to C-4 on the pyrrole. The Fe(IV)=O(OH_2) complex then abstracts a hydrogen atom from the C-4 of the pyrrole to form streptorubin B. Transfer of a second electron from the Rieske cluster to the non-haem iron centre and protonation

of the hydroxide ligand in the resulting Fe(II)(OH_2)(OH) complex, results in product release and regeneration of the reduced state of the enzyme to initiate another round of catalysis (Figure 1.14) (Barry and Challis, 2013).

To investigate the mechanism of the RedG-catalysed oxidative carbocyclisation of undecylprodigiosin (**2**) to streptorubin B (**3**), *Streptomyces albus* expressing *redHG* were fed with stereoselectively deuterium labelled 2-undecylpyrroles, [7′-2H](7′R)-2-undecylpyrrole and [7′-2H] (7′S)-2-undecyl-pyrrole, together with MBC (**16**). Monitoring deuterium incorporation showed that RedG abstracts predominantly the *pro*-R hydrogen atom from C-7′ of **2** (Withall *et al*., 2015). This result, together with the previous observation that **3** produced by *S. coelicolor* has the 7′S configuration, shows that the RedG-catalysed C-C bond forming reaction proceeds with inversion of configuration at C-7′. In contrast, in other non-haem iron-dependent enzymes such as isopenicillin N synthase and clavaminate synthase which catalyse oxidative cyclisation reactions, the step of carbon-heteroatom bond formation proceeds with retention of configuration at the carbon atom undergoing functionalisation. This stereochemical result is observed because the heteroatom participating in oxidative cyclisation is *cis* to the oxygen binding site of the non-haem iron centre in these enzymes (Withall *et al*., 2015).

Illumination of the mechanism of the RedG-catalysed oxidative carbocyclisation of **2** to **3** paves the way for understanding and investigating the mechanism and stereochemistry of similar oxidative carbocyclisation reactions catalysed by RedG homologues in the biosynthesis of other prodiginines.

1.13 Metacycloprodigiosin biosynthesis

As discussed in section 1.7, the 12-membered carbocyclic ring of metacycloprodigiosin (**19**) from *S. longispororuber* is also proposed to be biosynthesised by oxidative cyclisation of undecylprodigiosin (**2**).

To investigate this proposal, *mcpG* which encodes a RedG orthologue in *S. longispororuber* was cloned, sequenced and expressed in a *redG* mutant of *S. coelicolor*, resulting in production of **19**. The result shows RedG and McpG catalyse regio-stereodivergent oxidative carbocyclisation reactions (Figure 1.15) (Sydor *et al*., 2011).

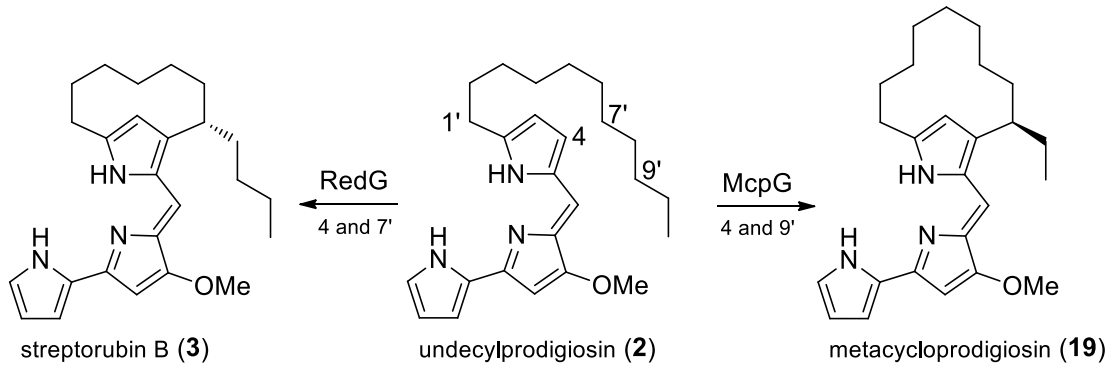


Figure 1.15 Regio-stereodivergent oxidative carbocyclisation reactions catalysed by RedG and McpG to form streptorubin B (**3**) and metacycloprodigiosin (**19**).

Based on comparison to the RedG mechanism, it can be proposed that the oxidative carbocyclisation of undecylprodigiosin to metacycloprodigiosin also proceeds with inversion of configuration, but involves abstraction of the *pro*-S hydrogen atom from C-9′. This prediction could be tested by conducting a deuterium labelled substrate experiment analogous to the one used to probe the mechanism of RedG.

1.14 Roseophilin and Prodigiosin R1 biosynthesis

Prodigiosin R1 (**20**) and roseophilin (**21**) are structurally related compounds containing two pyrrole and one furan ring produced by *Streptomyces grisioviridis*, which also belong to the prodiginine family (Figure 1.16) (Kawasaki *et al*., 2008; Hayakawa *et al*., 2009)

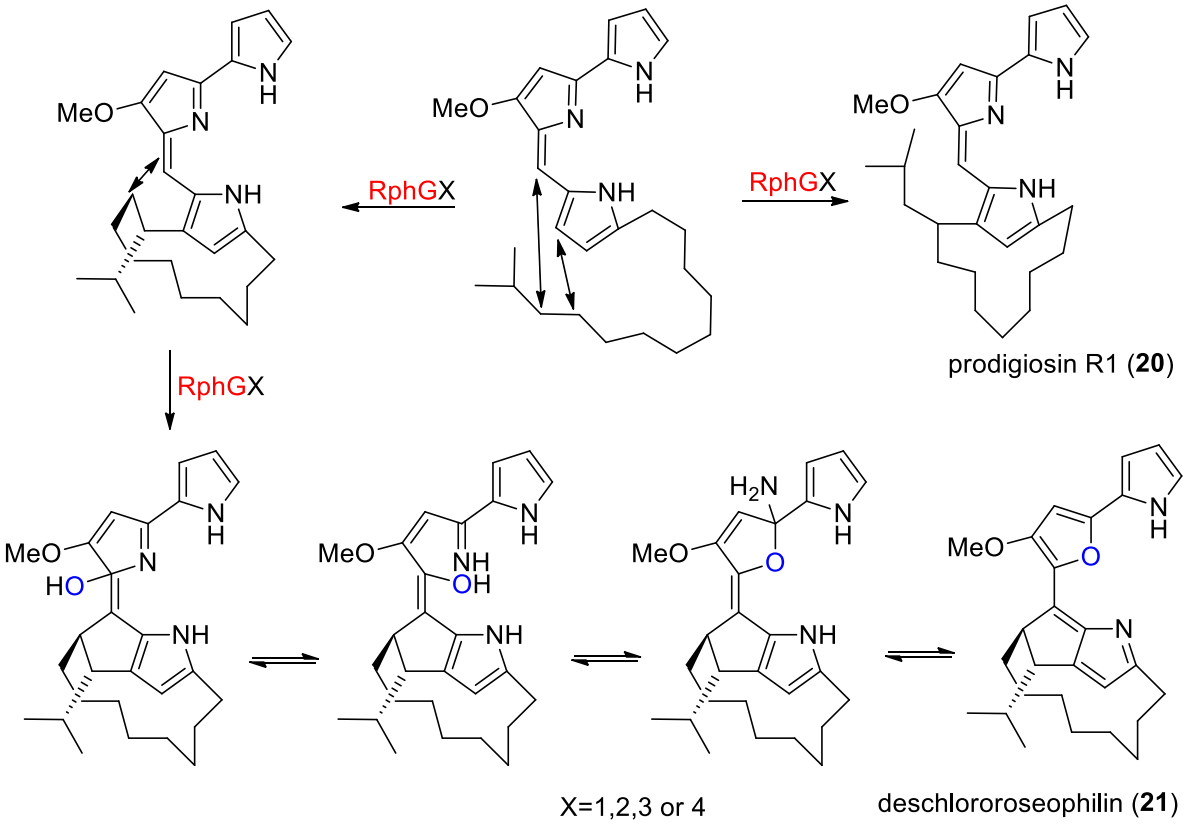


Figure 1.16 Proposed functions of RphG, RphG2, RphG3 and RphG4 in *Streptomyces griseoviridis*.

The *rph* cluster encoding for prodiginine biosynthesis in *S. griseoviridis* was characterised, cloned and sequenced (Kawasaki *et al*., 2009). The cluster contains twenty five genes or which twenty one are homologous to genes in the *red* cluster of *S.*

coelicolor including four putative oxygenase genes (*rphG*, *rphG2*, *rphG3* and *rphG4*) homologous to *redG*.

Sequence alignment of RphG 1-4 with RedG from *S. coelicolor* showed they are highly similar. They all contain conserved sequence motifs characteristics of a Rieske oxygenase like enzyme, except RphG3 which appears to be missing the N-terminal Rieske domain since it lacks the conserved Cys+His residues that ligate the [2Fe-2S] cluster. Presence of multiple oxygenase genes might explain the difference in number and positions of carbon-carbon bonds between **20** and **21,** but which enzyme catalyses which carbocyclisation reaction is not yet known.

1.15 Aims of the study

Recent research into the mechanism of action of RedG has given insight into how this remarkable enzyme is capable of catalysing oxidative carbocyclisation to give streptorubin B. To further understand this type of Rieske oxygenase this project will focus on characterisation of RedG homologues which carry out similar oxidative carbocyclisation reactions.

McpG, the Rieske oxygenase implicated in metacycloprodigiosin biosynthesis, has been shown to be able to complement a *S. coelicolor* strain lacking RedG, resulting in the formation of a cyclised product with different regio-and stereochemistry. The first aim was therefore to heterologously produce McpG to confirm its role in oxidative cyclisation of undecylprodigiosin. In order to further probe the mechanism of this reaction it was planned to develop a heterologous expression system for McpG and McpH. Such a system could then, in collaboration with Chris Perry (another PhD student in the Challis lab), be used to feed deuterium labelled undecylprodigiosin along with MBC to investigate the stereochemical course of the reaction. The role of McpH both in the coupling of MBC and 2-undecylpyrrole, and as a chaperone partner for McpG could also be investigated via these experiments.

The second aim of the project was to to investigate the function of RedG homologues proposed to be involved in roseophillin biosynthesis (RphG, RphG2, RphG3 and RphG4). The plan was to heterologously express each homologue in a suitable host to determine their ability to catalyse oxidative cyclisation of 11-methyldodecylprodigiosin.

Finally, as discussed in section 1.2, analysis of the genome sequence of *Streptomyces coelicolor* resulted in identification of a large number of novel natural product biosynthetic gene clusters which have subsequently been shown to direct the production

of several novel metabolites. Using bioinformatics analysis it has become increasingly possible to predict the likely products of many such gene clusters, particularly those that encode large multienzymes such as modular PKSs and NRPSs. Therefore, another aim of this project was to obtain the genome sequence of the metacycloprodigiosin producer *S. longispororuber*, to obtain the complete sequence of the metacycloprodigiosin biosynthetic gene cluster, along with several novel specialised metabolite biosynthetic gene clusters.

Results and Discussions

2. Investigation of *S. longispororuber redG* Orthologue: *mcpG*

2.1 Confirmation of metacycloprodigiosin production by *S. longispororuber*

As mentioned in chapter 1, metacycloprodigiosin (**19**) has previously been isolated from *S. longispororuber*. To confirm this, *S. longispororuber* was grown on R5 medium and its ability to produce **19** analysed. This extract was then used as authentic standard for **19** in later analysis. LC-MS showed that the strain produced undecylprodigiosin (**2**) and a carbocyclic derivative, presumed to be **19** (Figure 2.1).

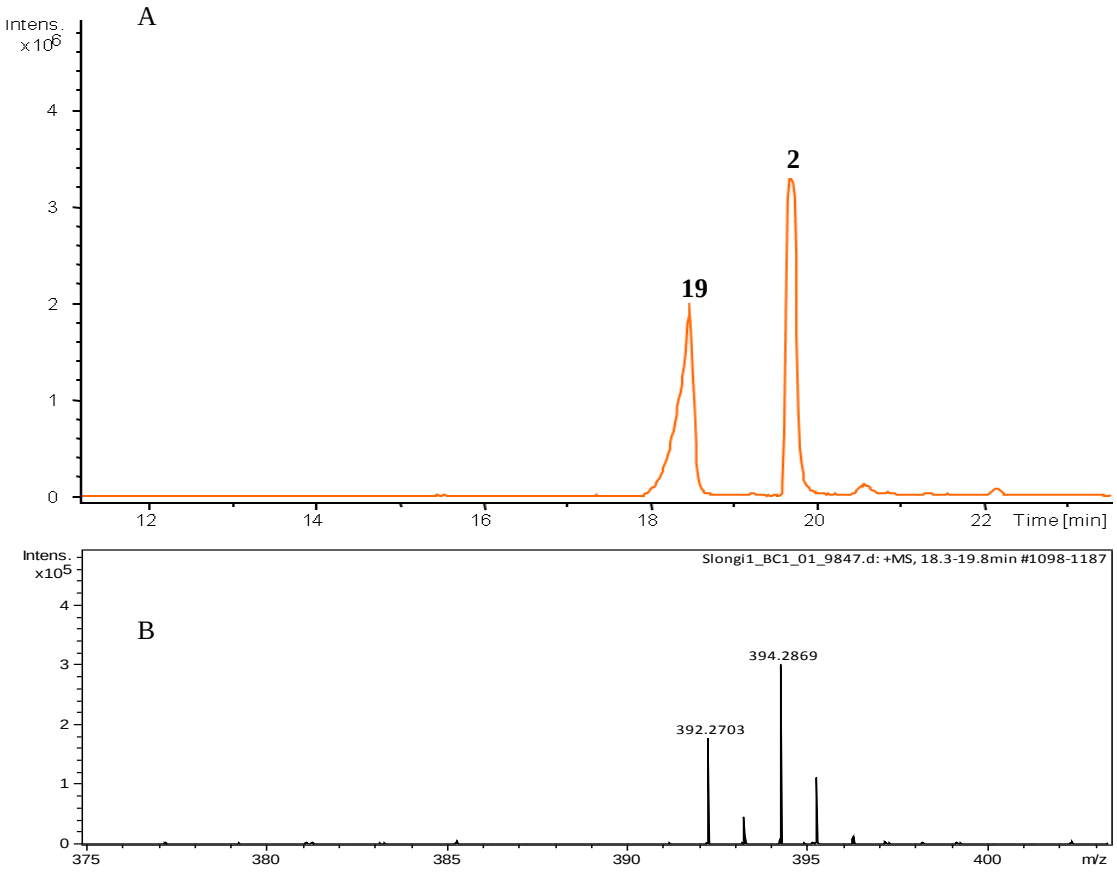


Figure 2.1 Metacycloprodigiosin production by *S. longispororuber*. A. EIC for *m/z* range 392-394 (corresponding to $[M+H]^+$ for undecylprodigiosin (**2**) and metacycloprodigiosin (**19**) from LC-MS analysis of acidified organic extract of *S. longispororuber*. B. HR-MS confirming the molecular formula of **19** (calculated for $C_{25}H_{35}N_3O$:392.2696, found: 392.2703) and for **2** (calculated for $C25H_{37}N_3O$: 394.2856, found: 394.2869).

2.2 Cloning strategy for heterologous expression

2.2.1 pOSV556

The vector that the targeted genes were cloned into was pOSV556 (Figure 2.2). The genes expressed from the vector are under the control of the constitutive *ermE* promoter. It contains oriT, which allows conjugal transfer of the vector from *E. coli* to *Streptomyces*. The integration of this vector into the chromosome of *Streptomyces* is catalysed by an integrase that is encoded within pOSV556. The integration can occur through site-specific recombination between the *attP* site within the vector and *attB* site within the chromosome. The integration is catalysed by an integrase within the vector (Raynal *et al*., 2002). The vector contains two antibiotic resistance genes: one for ampicillin (for selection in *E. coli*) and the other for hygromycin (for selection in *Streptomyces*). This vector was used to investigate the function of specific genes via heterologous expression in *Streptomyces* strains.

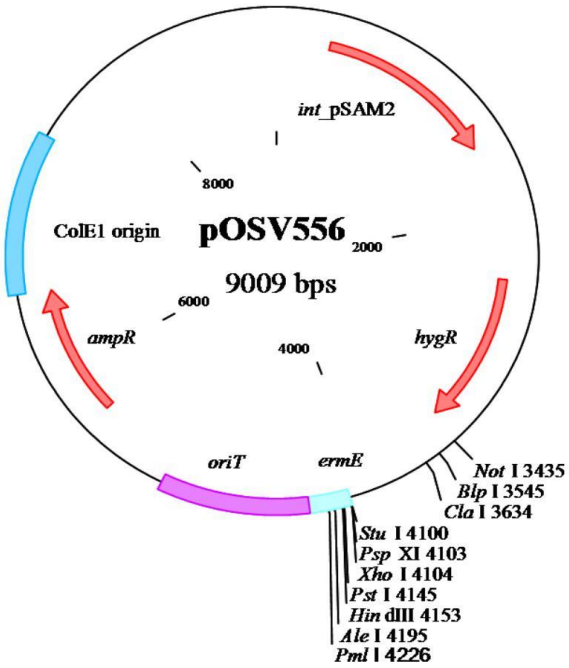


Figure 2.2 The pOSV556 vector.

2.2.2 Cloning strategy

PCR amplifications were carried out using a forward primer to anneal upstream of the start codon of the gene of interest including the natural ribosomal binding site (RBS). In some experiments, an artificial conserved RBS (5'-AGGAGG-3') (Strohl, 1992) was introduced into the forward primers. Reverse primers were designed to anneal downstream of the stop codon of the gene. *Hin*dIII restriction site was used within the forward primers and *Stu*I, *Xho*I or *Pst*I restriction sites were used within the reverse primers. The PCR amplified products were introduced into pOSV556 in its multiple cloning site using appropriate restrictions enzymes. Digestion and ligation of the PCR product and the pOSV556 vector were performed using standard cloning procedures.

The pOSV556 construct was then used to transform *E. coli* TOP10 by heat-shock or electroporation and growing colonies were selected on LB agar containing ampicillin. Correct insertion of the PCR product into the vector was confirmed by restriction digestion, PCR amplification and sequencing. Plasmid clones were transferred to *E. coli* ET12567/pUZ8002 via electroporation and then from *E. coli* ET12567/pUZ8002 to the appropriate *Streptomyces* strain via conjugation. Few hygromycin resistant transconjugants were chosen and the conjugation success was confirmed by PCR amplification.

2.3 *Streptomyces* hosts for heterologous expression

Protein sequence comparison of McpG and RedG by Challis group showed high similarity including the presence of the conserved sequence motifs characteristic of Rieske oxygenase-like enzymes (Sydor, 2010). To confirm the role of McpG in the final step of prodiginine biosynthesis, we aimed to express it in different *Streptomyces*

hosts such as *Streptomyces coelicolor* M511 and *Streptomyces albus. S. coelicolor* is an ideal model strain for the study of the biosynthesis of specialised metabolites such as the prodiginines as it lacks plasmids SCP1 and SCP2 and the actinorhodin pathway specific activator gene (*actII–ORF4*) (Floriano and Bibb, 1996) and *S. albus* does not produce any prodiginine like metabolites (Traxler *et al*., 2013).

2.4 Heterologous expression of *S. longispororuber redG* orthologue: *mcpG in S. albus* and *S. coelicolor* M511

Both pOSV556/*redG* and pOSV556/*mcpG* (from the lab collection) were purified and analysed by restriction digestion using *Xho*I which cuts inside *redG* and *mcpG* as well as a common region downstream of the 5` of the ends of the genes giving restriction fragment of 956bp and 724bp respectively (Figure 2.3). Following this, *E. coli* ET12567/pUZ8002 was transformed separately with each plasmid using electroporation.

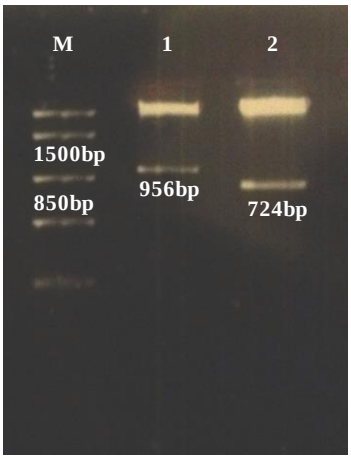


Figure 2.3 Agarose gel electrophoretic analysis of *Xho*I restriction digests of pOSV556*redG* (lane 1) and pOSV556*mcpG* (lane 2). The lane marked M contains molecular size markers.

Each of the constructs pOSV556*redG* and pOSV556*mcpG* were separately transferred from *E. coli* ET12567/pUZ8002 into *Streptomyces* (*S. coelicolor* M511 and *S. albus*)

via conjugation, and hygromycin resistant colonies were selected. These transconjugant *Streptomyces* strains were grown on R5 medium on a semi-permeable membrane and fed with undecylprodigiosin (**2**) extracted from *S. coelicolor* W31 (M511Δ*redG*::*scar*; which produces **2** but not streptorubin B (**3**). After 3 days the cells were extracted with acidified methanol (Section 5.2.13) and the extracted products were analysed by LC-MS (Figure 2.4 and 2.5). No streptorubin B (**3**) or metacycloprodigiosin (**19**) was produced by *S. coelicolor* M511/pOSV556*redG* (B) and *S. coelicolor* M511/pOSV556*mcpG* (C) respectively (Figure 2.4).

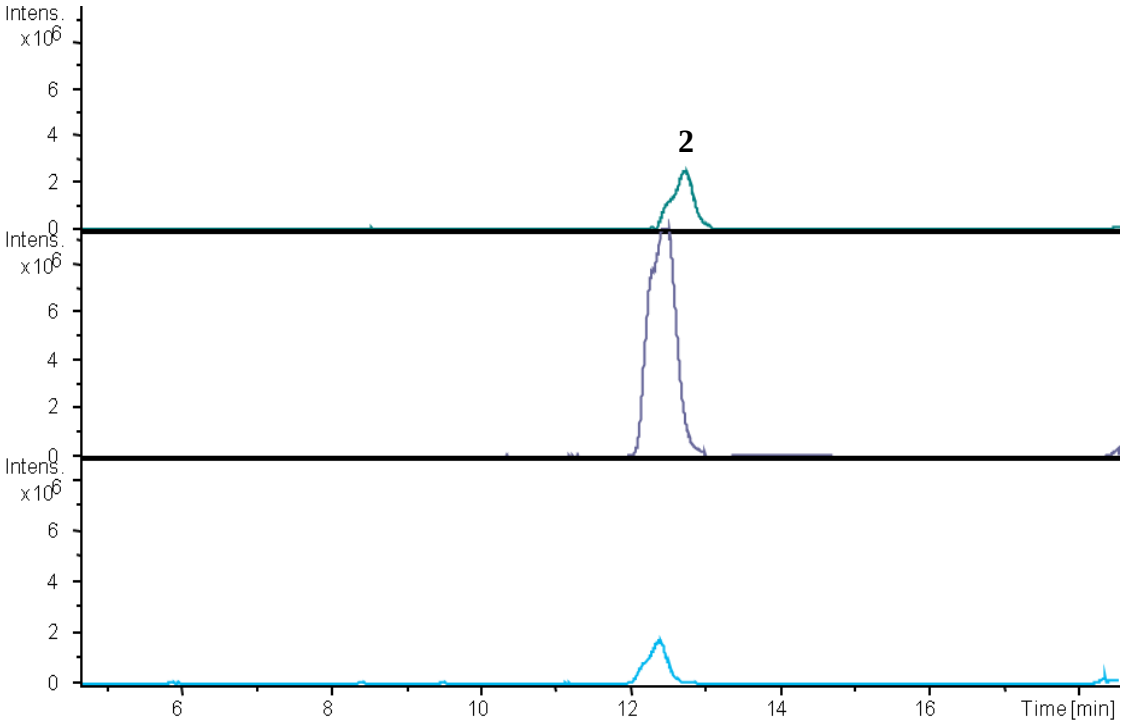


Figure 2.4 Activity assay of McpG and RedG in *S. coelicolor* M511. EIC for *m/z* range 392-394 (corresponding to $[M+H]^+$ for undecylprodigiosin (**2**) and metacycloprodigiosin (**19**) from LC-MS analysis of acidified organic extracts of *S. coelicolor* M511 (top trace), *S. coelicolor* M511/pOSV556*redG* (middle trace) and *S. coelicolor* M511/pOSV556*mcpG* (bottom trace) fed with undecylprodigiosin extracted from *S. coelicolor* W31(Δ*redG*).

However, a small amount of cyclisation to form **3** and **19** could be observed in the *S. albus* transconjugants expressing *redG* (B) and *mcpG* (C), respectively, compared to

wild type *S. albus* (A) (Figure 2.5). It was, therefore concluded that *S. albus* was a better host for heterologous expression.

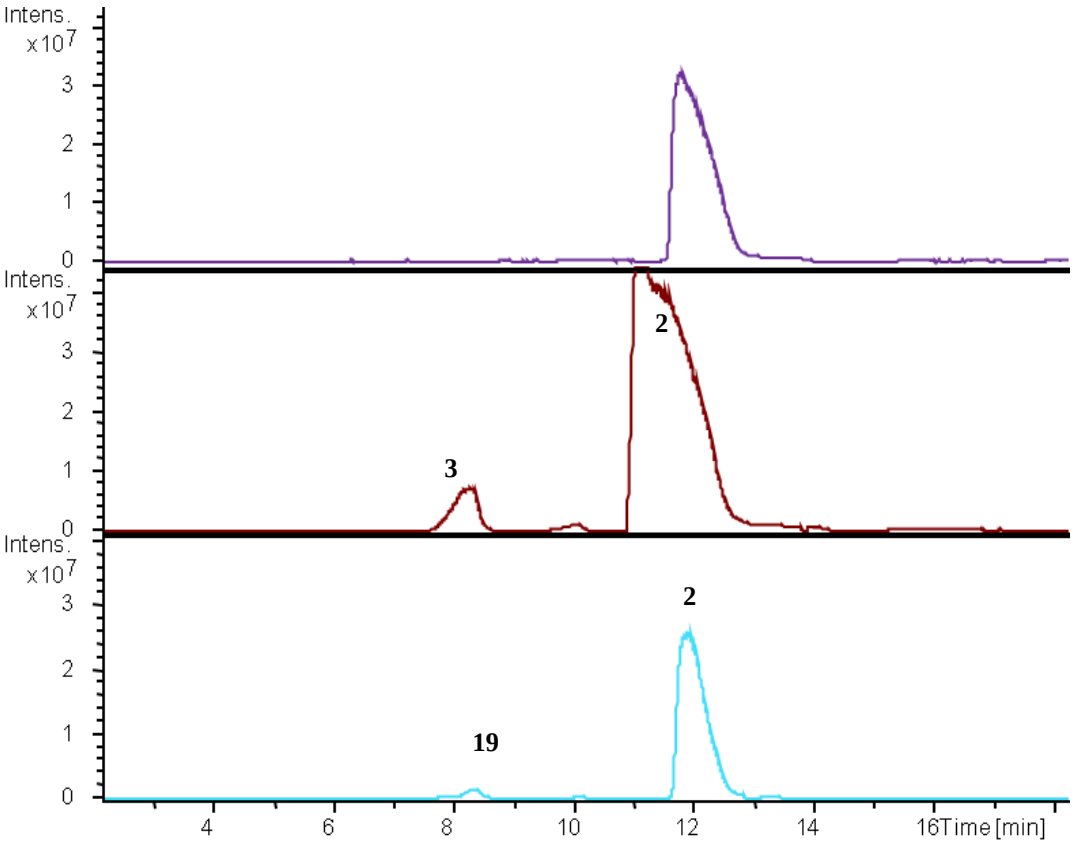


Figure 2.5 Activity assay of McpG and RedG in *S. albus*. EIC for *m/z* range 392-394 (corresponding to $[M+H]^+$ for undecylprodigiosin (**2**), streptorubin B (**3**) and metacycloprodigiosin (**19**) from LC-MS analyssi of acidified organic extracts of *S. albus* (top trace), *S. albus*/pOSV556*redG* (middle trace) and *S. albus*/pOSV556*mcpG* (bottom trace) fed with undecylprodigiosin extracted from *S. coelicolor* W31(Δ*redG*).

The low level of metacycloprodigiosin (**19**) produced by *S. albus* expressing *mcpG* and streptorubin B (**3)** produced by *S. albus* expressing *redG* could be due to the absence of McpH and RedH which is proposed to form a complex with McpG and RedG (Sydor, 2010). To investigate this possibility we aimed to coexpress *mcpH* and *mcpG*.

2.5 Heterologous expression of *S. longispororuber mcpH*

The role of *mcpH* in prodiginine biosynthesis was first investigated to confirm that it has the same function as *redH* (the condensation of 2-UP (**15**) and MBC (**16**) to form undecylprodigiosin (**2**)). Protein sequence comparison of RedH from *S. coelicolor* and McpH from *S. longispororuber* showed very high similarity (78%).

Clones of *mcpH* and *mcpHG* (from *S. longispororuber*, from the lab collection) in pOSV556 were analysed by restriction digestion (*Bam*HI) (Figure 2.6) and sequencing to check their integrity. The digestion results indicated a part of the vector may have been lost, since, although the control reaction using pOSV556 gave the expected three bands, neither pOSV556*mcpH* or pOSV556*mcpHG* were digested at all (Figure 2.6).

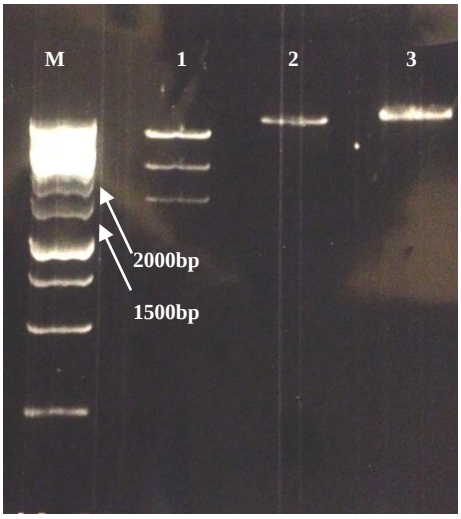


Figure 2.6 Agarose gel electrophoretic analysis of pOSV556 (lane 1), pOSV556/*mcpH* (lane 2) and pOSV556/*mcpHG* (lane 3) digested with *BamH*I. The lane marked M contains1kb molecular size markers

2.5.1 Cloning of *S. longispororuber mcpH*

To repeat the cloning, a new set of PCR primers were designed for amplification of *mcpH* (Table 5.2). The *mcpH* gene, including its natural ribosomal binding site (RBS) AGGAGG 7bp upstream of the start codon, was amplified from the 3G3 fosmid created by Paulina Sydor (Sydor, 2010).

mcpH (2835bp) was amplified using PCR (Figure 2.7), and the band with the correct molecular size was purified following agarose gel electrophoresis. The amplified gene and pOSV556 were both digested with the restriction enzymes *Hin*dIII and *Stu*I. The restriction products were then isolated using agarose gel electrophoresis and ligated together.

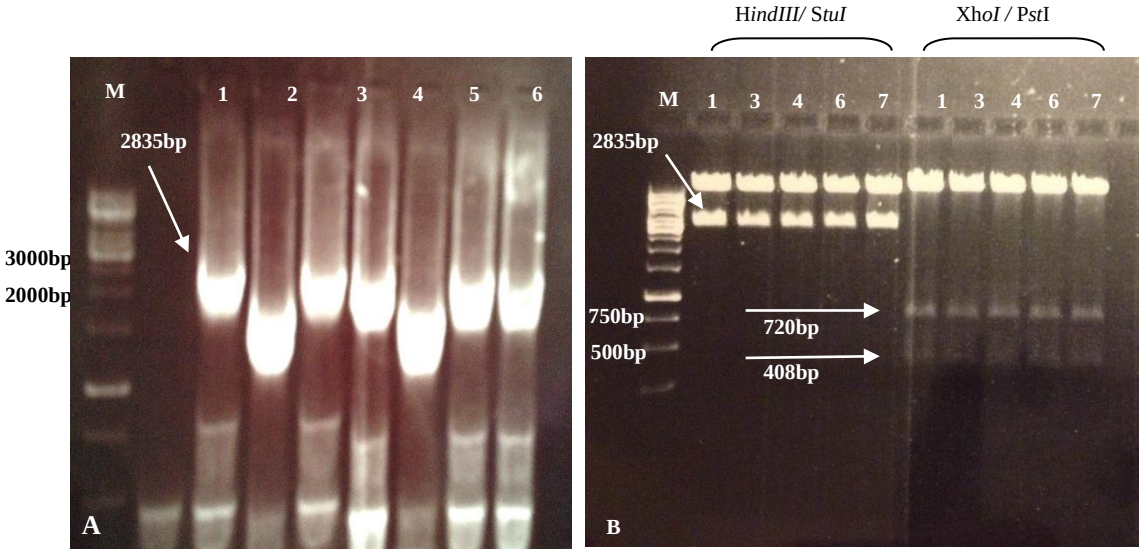


Figure 2.7 Agarose gel electrophoretic analysis of A. PCR amplification of *mcpH*. B. Digestion of pOSV556/*mcpH* with *Hin*dIII/*Stu*I and *Xho*I/ *Pst*I. The lanes marked M contains1kb molecular size markers Expected band sizes are labelled.

The product of the ligation reaction was purified and used to transform *E. coli* TOP10 competent cells, selecting for ampicillin resistance. PCR was used to identify positive clones (Figure 2.7). Five clones appeared to contain *mcpH* and were further analysed by restriction digestion with *Hin*dIII/*Stu*I and *Xho*I/*Pst*I (Figure 2.7). All four gave

restriction products of the expected sizes, confirming they contained the correct insert. Two clones were sequenced and shown to be correct.

To investigate the role of McpH in condensation of 2-UP (**15**) and MBC (**16**), *S. albus* was chosen for heterologous expression because it does not contain any gene cluster similar to the *mcp* cluster. Therefore it contains no genes similar to *mcpH* and it does not produce prodiginines. Furthermore, it has been shown to be a better host than *S. coelicolor* for expression of *mcpG*. Purified pOSV556t/*mcpH* was used to transform *E. coli* ET12567/pUZ8002 using electroporation. The construct pOSV556/*mcpH* was then transferred from *E. coli* ET12567/pUZ8002 into *S. albus* via conjugation, and hygromycin resistant colonies were selected.

2.5.2 Feeding with 2-UP and MBC

The transconjugant *S. albus* cells expressing *mcpH* were grown on R5 medium on a semi-permeable membrane and fed with 2-UP (**15**) and MBC (**16**). The extracted products were analysed by LC-MS (Figure 2.8).

LC-MS analysis of acidified organic extracts showed the presence of large quantities of red-pigment undecylprodigiosin (**2)** in *S. albus* expressing *mcpH* with the same retention time as **2** produced by *S. longispororuber*. The molecular formula of **2** was confirmed by HR-MS and **19** was not observed. In extracts of wild type *S. albus* neither compound was observed. This confirms the role of McpH as an enzyme that catalyses the condensation of **15** and **16** to form **2** in *S. longispororuber*.

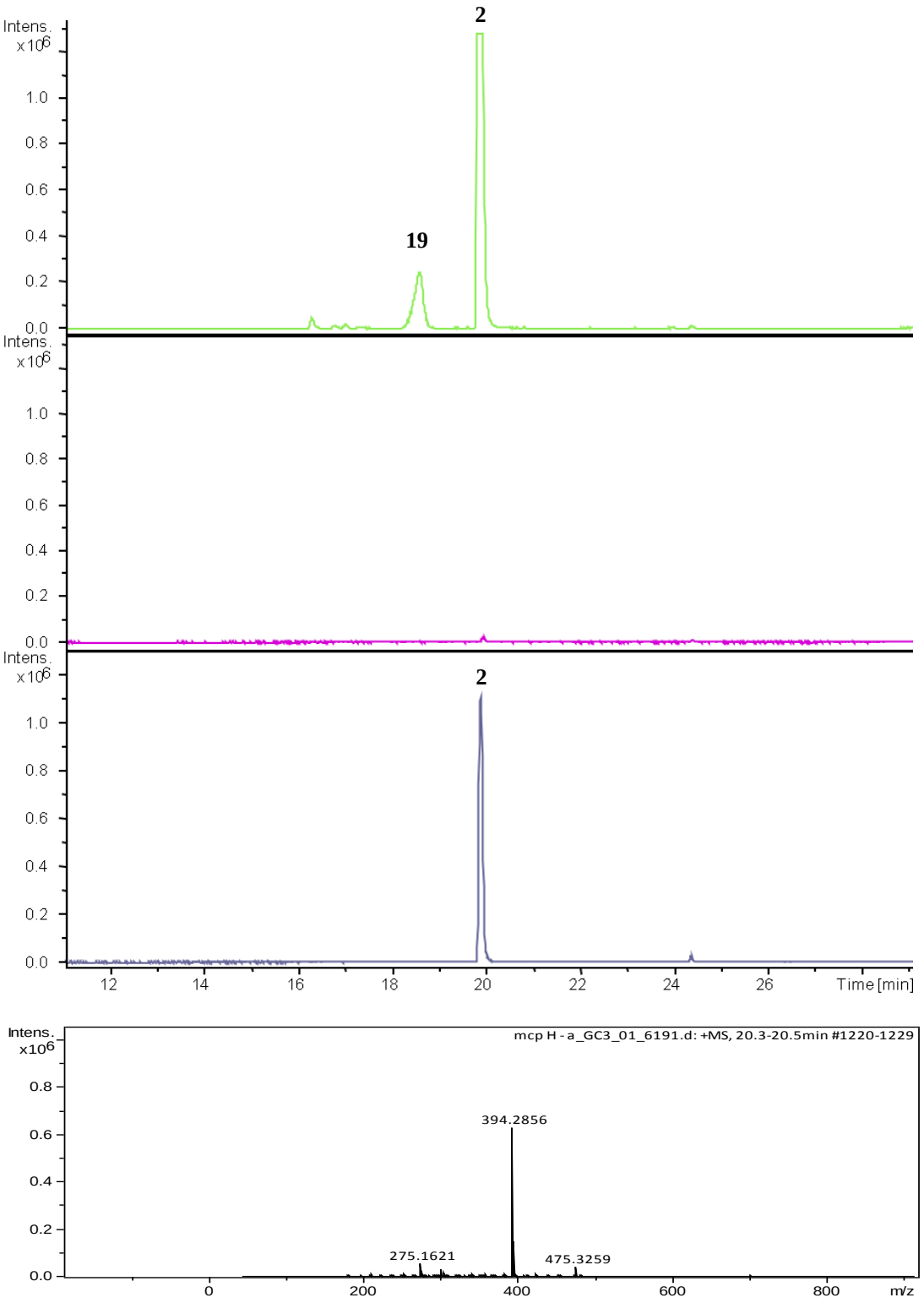


Figure 2.8 Activity assay for McpH in *S. albus*. A. EIC for *m/z* range 392-394 (corresponding to $[M+H]^+$ for undecylprodigiosin (**2**) and metacycloprodigiosin (**19**) from LC-MS analysis of acidified organic extracts of *S. longispororuber* (top trace), *S. albus* (middle trace) and *S. albus*/pOSV556/*mcpH* (bottom trace) fed with 2-UP (**15**) and MBC (**16**). B. HR-MS confirming the molecular formula of **2** (calculated for $C_{25}H_{37}N_3O$: 394.2856, found; 394.2856).

2.6 Heterologous expression of *S. longispororuber redHG* Orthologues: *mcpHG*

To overcome the problem of low levels of production of metacycloprodigiosin (**19**) when expressing *mcpG* alone in *S. albus*, a heterologous expression system in which both *mcpH* and *mcpG* were expressed, was prepared to investigate whether formation of a complex between the two enzymes is required for full functionality.

2.6.1 Cloning of *S. longispororuber redHG* Orthologues: *mcpHG*

Following amplification of *mcpHG*, agarose gel electrophoresis was used to analyse the products (Figure 2.9), and the band with the correct molecular weight was purified from the gel. The PCR product and pOSV556 were digested with the restriction enzymes *Hin*dIII and *Stu*I, the vector and insert were ligated together and the product of the ligation was purified and used to transform *E. coli* TOP10 competent cells selecting for ampicillin resistance. PCR and restriction digestion were used to analyse the resulting clones, showing that the ligation reaction had been successful. Two clones were then sequenced. The construct pOSV556/*mcpHG* was transferred from *E. coli* ET12567/pUZ8002 into *S. albus* via conjugation, and hygromycin resistant colonies were selected.

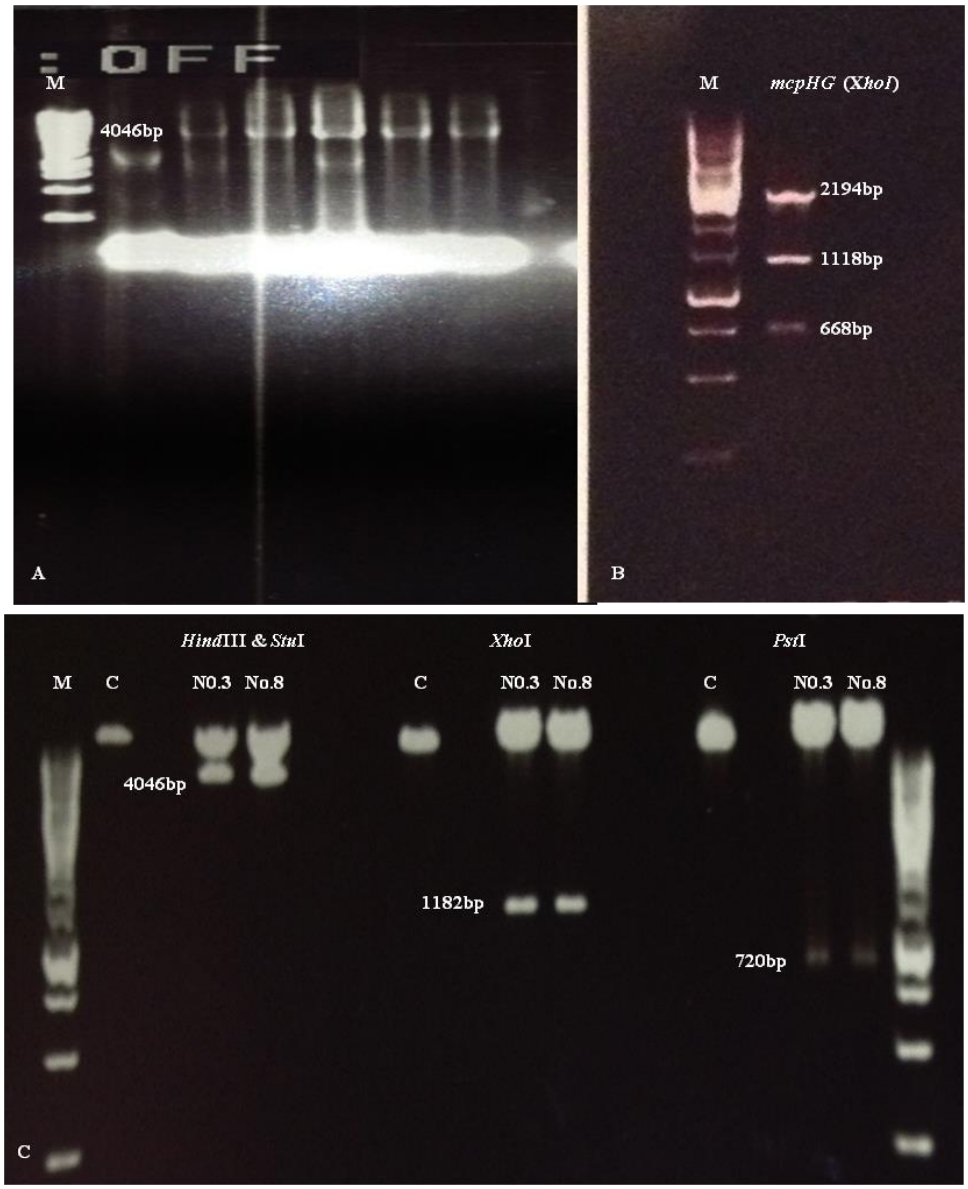


Figure 2.9 Agarose gel electrophoretic analysis, A. PCR of *mcpHG* (3927bp). B. Restriction digestion of *mcpHG* PCR products with *Xho*I, C. Restriction digestion of pOSV556 (Control), pOSV556/*mcpHG* clone no.3, (2) and pOSV556/*mcpHG* clone no. 8 with *Hin*dIII and *Stu*I, *Xho*I. The lane marked M contains 1kb molecular weight markers. Expected band sizes are labelled in white.

2.6.2 Feeding with 2-UP and MBC

S. albus, *S. albus*/*mcpH* and *S. albus*/*mcpHG* were grown on R5 medium on a semi-permeable membrane and fed with 2-UP (**15**) and MBC (**16**) synthesised by David

Withall (a former PhD student) and Chris Perry (a current PhD student) in the Challis group.

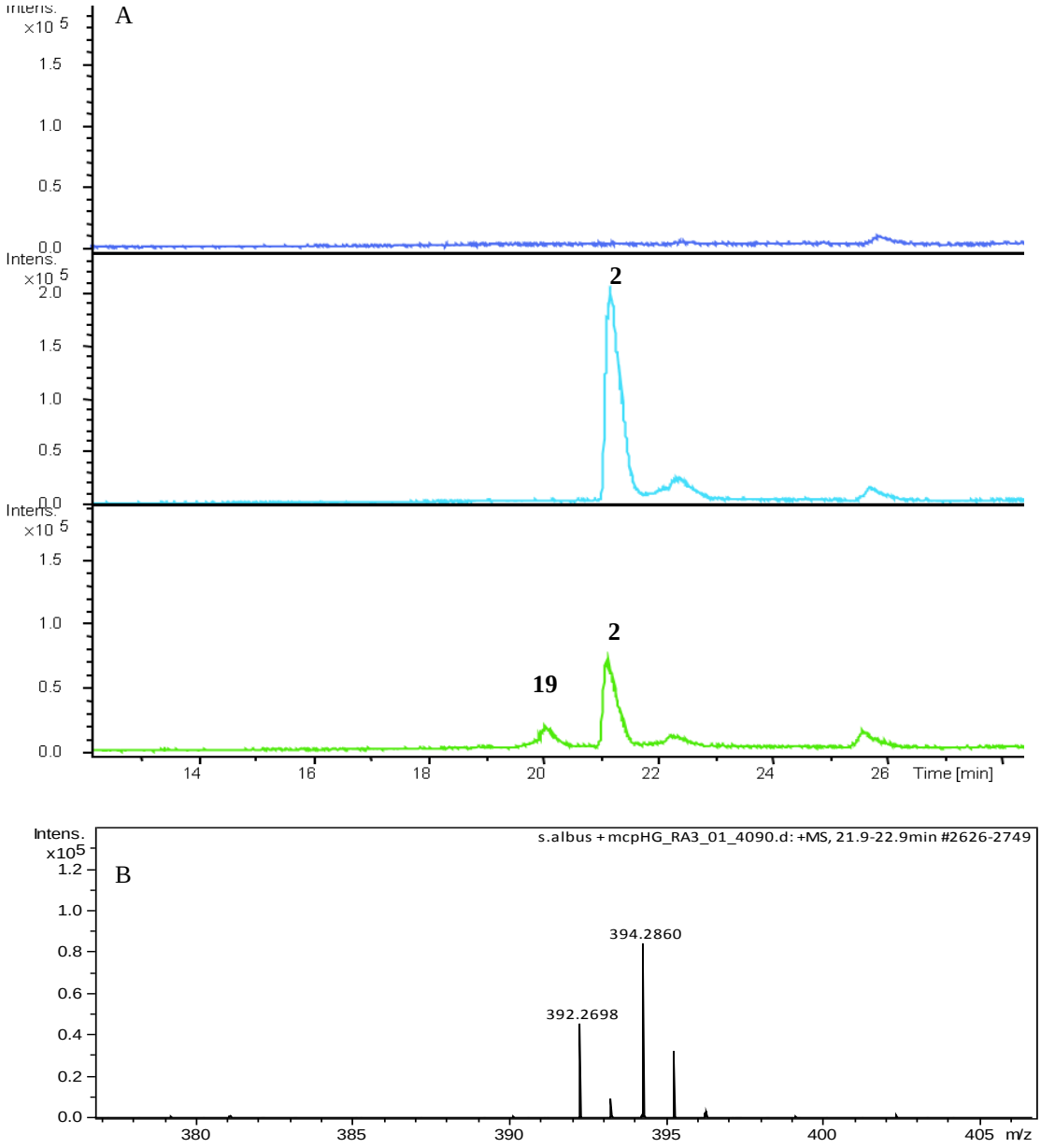


Figure 2.10 Activity assay of McpHG in *S. albus*. A. EIC for *m/z* range 392-394 (corresponding to $[M+H]^+$ for undecylprodigiosin (**2**) and metacycloprodigiosin (**19**) from LC-MS analysis of acidified organic extracts of *S. albus* (top trace), *S. albus*/pOSV556/*mcpH* (middle trace) and *S. albus*/pOSV556/*mcpHG* (bottom trace) fed with 2-UP (**15**) and MBC (**16**). B. HR-MS confirming the molecular formula of **19** (calculated for $C_{25}H_{35}N_3O$: 392.2696, found: 392.2698) and for **2** (calculated for $C_{25}H_{37}N_3O$: 394.2856, found: 394.2860).

The extracted products were analysed by LC-MS, which revealed that in the *S. albus/mcpHG* strain both undecycloprodigiosin (**2**) and a cyclic derivative, assumed to be metacycloprodigiosin (**19**), were produced whereas there was only (**2**) in *S. albus*/*mcpH* and neither **2** nor **19** were produced by wild type *S. albus* (Figure 2.10). The amount of the carbocyclised metabolite was higher than when the experiment was conducted with *S.albus* expressing *mcpG* alone (Section 2.4, Figure 2.5), but this was still not sufficient for NMR spectroscopic characterisation.

2.7 Re-examination of the sequence of *mcpG*

Due to the problems encountered with low levels of formation of **19**, the sequence of *mcpG* was re-examined to ensure there were no inaccuracies. Sequence comparison of *mcpG* from 3G3 fosmid template (prepared by Paulina Sydor) with *redG* from *S. coelicolor* showed that *mcpG* construct was 48 amino acids shorter than *redG* due to a C-terminal truncation (Sydor, 2010).

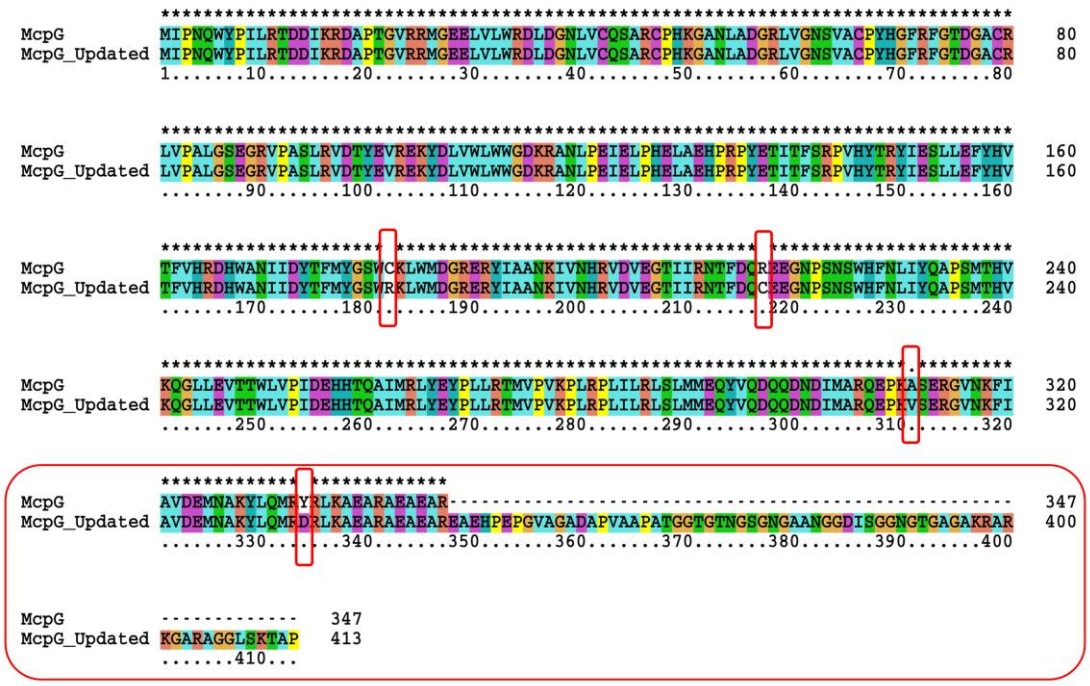


Figure 2.11 Protein sequence alignment of the McpG and McpG (updated) from *S. longispororuber*.

To be completely certain of the correct sequence of *mcpG* the genome of *S. longispororuber* was sequenced (Figure 2.11) (see chapter 4). The new sequence data revealed an error in the original *mcpG* sequence, resulting in the omission of 66 amino acids from the C-terminus of the protein. To determine whether it was truncation of the enzyme that caused low McpG activity, *mcpHG* was recloned into pOSV556 and expressed in *S. albus*.

2.7.1 Expression of corrected *mcpHG* in *S. albus* and feeding with 2-UP and MBC

Cloning, confirmation and conjugation (Figure 2.12) were performed using the procedures described in Section 5.2 and discussed throughout this chapter.

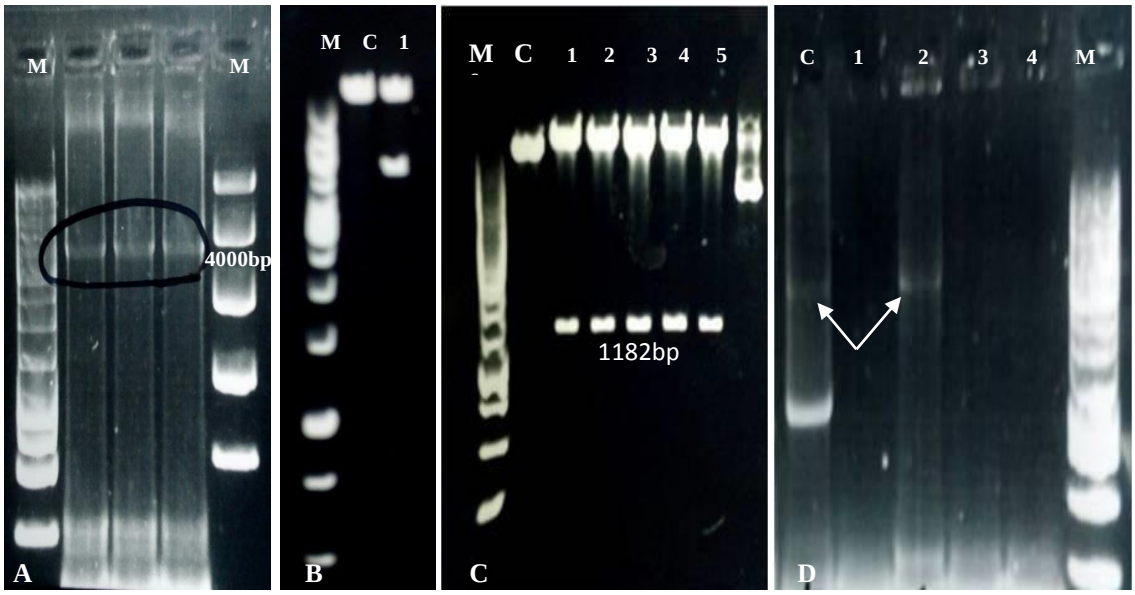


Figure 2.12 Agarose gel electrophoretic analysis, A. PCR amplification of *mcpHG*. B. pOSV556/*mcpHG* digested with *Hin*dIII and *Stu*I. C. pOSV556/*mcpHG* digested with *Xho*I. D. PCR amplification of *mcpHG* from transconjugant *S. albus*. The lane marked M contains 1kb molecular weight markers. Expected band sizes (3981bp) in are labelled in white.

S. albus, and *S. albus/mcpHG* were grown on R5 medium on a semi-permeable membrane and fed with 2-UP (**15**) and MBC (**16**). LC-MS analysis, following methanol extraction of *S. albus* showed although production of metacycloprodigiosin

(**19**) could be observed the amount of conversion to cyclised product was very low (Figure 2.13).

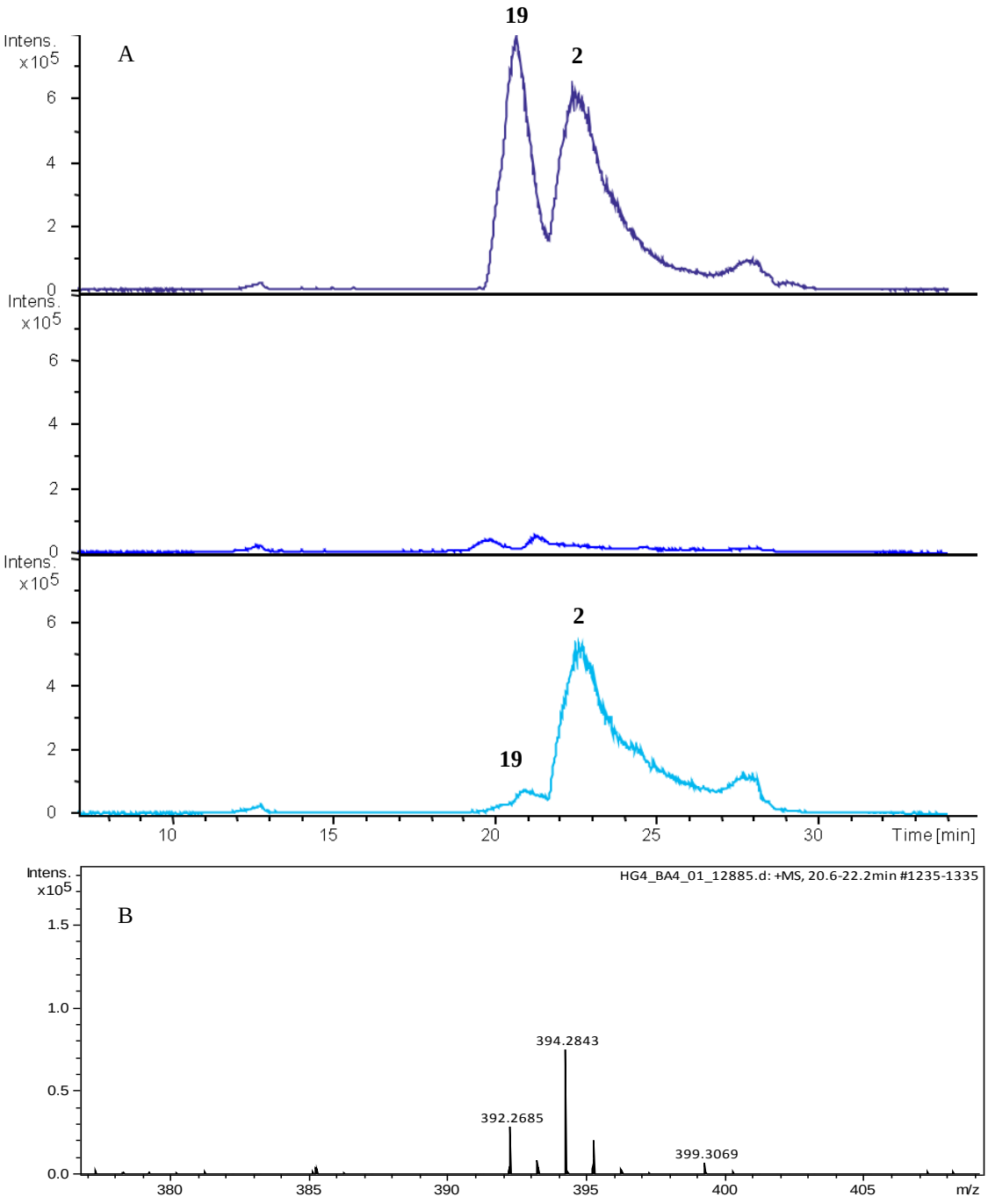


Figure 2.13 Activity assay of McpH in *S. albus*. A. EIC for *m/z* range 392-394 (corresponding to $[M+H]^+$ for undecylprodigiosin (**2**) and metacycloprodigiosin (**19**) from LC-MS analysis of acidified organic extracts of *S. longispororuber* (top trace), *S. albus* (middle trace) and *S. albus*/pOSV556*mcpHG* (bottom trace) fed with 2-UP (**15**) and MBC (**16**). B. HR-MS confirming the molecular formula of **19** (calculated for $C_{25}H_{35}N_3O$: 392.2696, found: 392.2685) and for **2** (calculated for $C_{25}H_{37}N_3O$: 394.2856, found: 394.2843).

A possible explanation for the low activity of McpG in these experiments is that it releases hydrogen peroxide, which is toxic to the cell, in the absence of its substrate, resulting in a strong selective pressure for less active variants. Therefore, expression of the corrected *mcpG* alone in *S. coelicolor redG* mutant was conducted.

2.7.2 Expression of corrected *mcpG* in *S. coelicolor* W31 (M511Δ*redG*::*scar*)

As mentioned in section 1, the role of McpG in the oxidative cyclisation of undecylprodigiosin (**2**) to give metacycloprodigiosin (**19**) was confirmed by heterologous expression of *mcpG* in *S. coelicolor* W31 (M511Δ*redG*::*scar*). This resulted in conversion of the **2** produced by this strain to **19**. The McpG sequencing error discovered in this work means that the enzyme used for this experiment was a truncated variant of the full length enzyme. It was therefore necessary to repeat the experiment with the full length construct. The gene was amplified using PCR primers as described (chapter 5) with a *Hin*dIII restriction site at the 5' end of the forward primer and a *Stu*I site at the 5' end of the reverse primer. The amplified PCR product was cloned into pOSV556 and the integrity of the gene was confirmed (Figure 2.14).

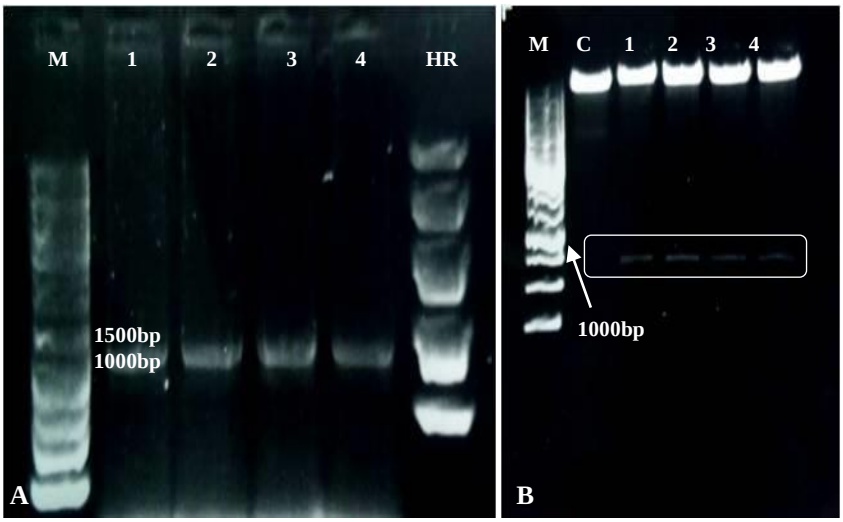


Figure 2.14 Agarose gel electrophoretic analysis, A. PCR amplification of *mcpG*. B. pOSV556/*mcpG* digested with *Hin*dIII and *Xho*I. The lane marked M contains 1kb molecular markers and the HR contains high range molecular weight markers.

The transformed cells were grown on LB agar containing ampicillin and correct clones were identified. The construct was introduced into *E. coli* ET12567/PUZ8002 and then transferred to *S. coelicolor* W31 via conjugation and hygromycin resistant transconjugants were selected to be grown on R5 medium.

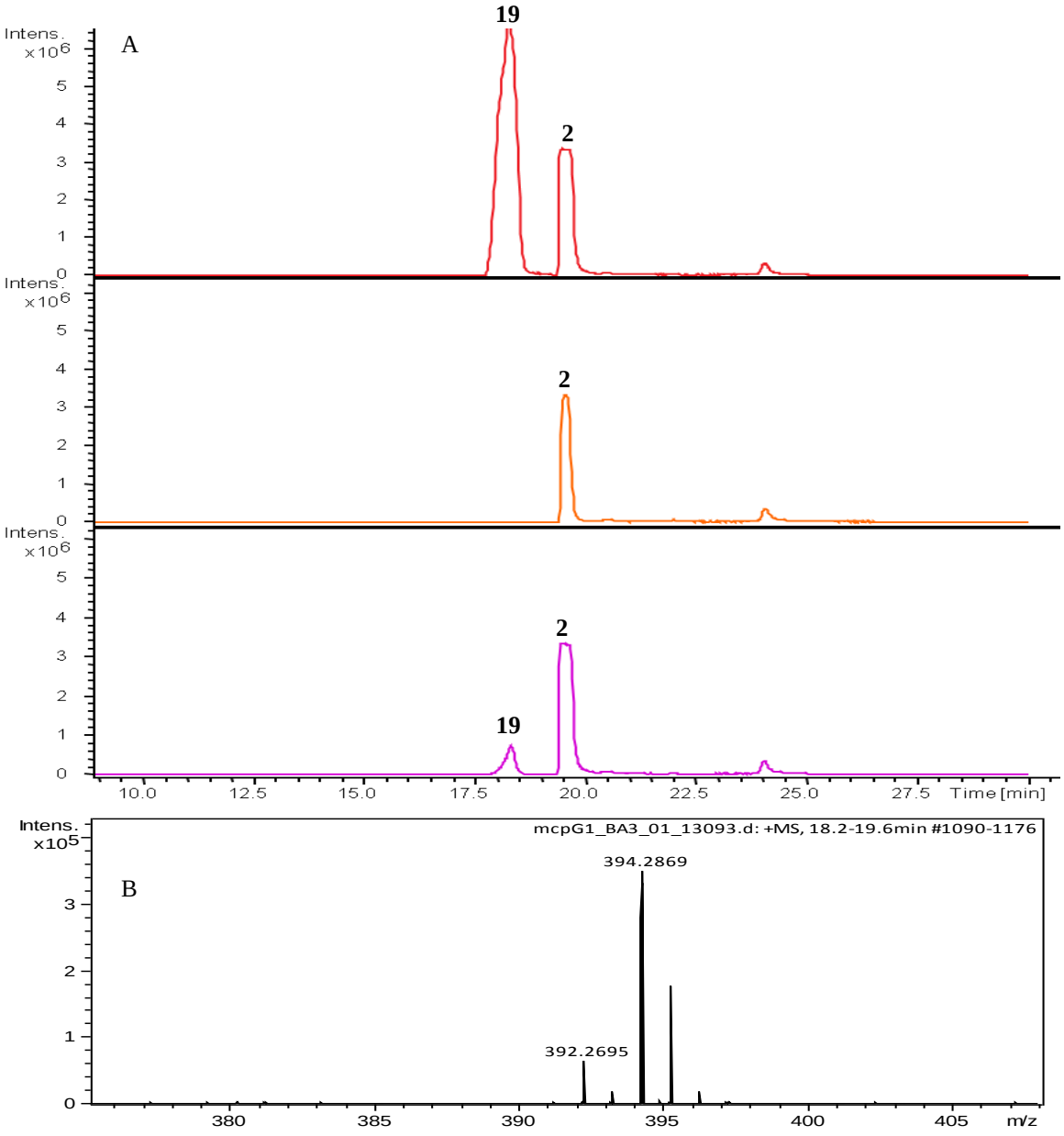


Figure 2.15 Metacycloprodigiosin production by *S. longispororuber*. A. EIC for *m/z* range 392-394 (corresponding to $[M+H]^+$ for undecylprodigiosin (**2**) and metacycloprodigiosin (**19**) from LC-MS analysis of acidified organic extract of *S. longispororuber* (top trace), *S. coelicolor* W31 (middle trace) and *S. coelicolor* W31/pOSV556*mcpG*. B. HR-MS confirming the molecular formula of **19** (calculated for $C_{25}H_{35}N_3O$: 392.2696, found: 392.2695) and **2** (calculated for $C_{25}H_{37}N_3O$: 394.2856, found: 394.2869).

LC-MS analysis of the acidified organic extracts showed two peaks with masses corresponding to undecylprodigiosin (**2**) (*m/z* = 394.2974) and a carbocyclic derivative (*m/z* = 392.2624) (Figure 2.15).

S. coelicolor W31 M511/*mcpG* was grown on R5 medium for one week and the cyclised metabolite with *m/z* =392.2624 was purified from large scale cultures using preparative HPLC. The molecular formula of this compound was confirmed by HR-MS. Since streptorubin B (**3**) and metacycloprodigiosin (**19)** have the same retention time, to confirm the identity of the purified metabolite, it was analysed by ^{1}H NMR and ^{13}C NMR spectroscopy. **3** can be distinguished from **19** by the presence of a very unusual and characteristic signal at -1.53 ppm for one of the C'4 hydrogen atoms. The corresponding signal in the ^{1}H NMR spectrum of **19** has a chemical shift of 0.2 ppm (Figure 2.17A).

Table 2.1 ^{13}C NMR data for metacycloprodigiosin in $CDCl_3$. A. metacycloprodigiosin from *S. longisporusruber*. B. metacycloprodigiosin from *redG* mutant of *S. coelicolor* expressing *mcpG*.

No.	A	B
1	126.8	126.2
2	111.6	111.6
3	116.8	117.4
4	122.3	122/2
5	147.4	148.8
6	92.7	93.2
7	165.7	166
8	120.6	120.6
9	113.3	113.4
10	125.9	126.2
11	150.3	150.5
12	112.4	112.5
13	154.2	154.5
14	28.9	29
15–21	34.4	34.46
	27.3	27.46
	26.8	26.8
	26.6	26.69
	25.5	25.64
	24.5	24.6
	22.8	22.89
22	39.6	39.8
23	29.9	29.8
24	12.6	12.6
25	58.7	58.8

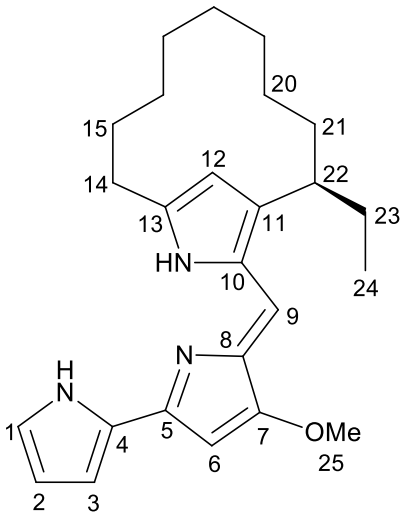


Figure 2.16 NMR analysis of metacycloprodigiosin (19).

Furthermore, methine protons and hydrocarbon protons α to the pyrrole ring have characteristically different shifts. ^{13}C NMR spectroscopic analysis unambiguously confirmed the cyclic metabolite to be metacycloprodigiosin (**19**) (Table 2.1, Figure 2.16 and Figure 2.17B).

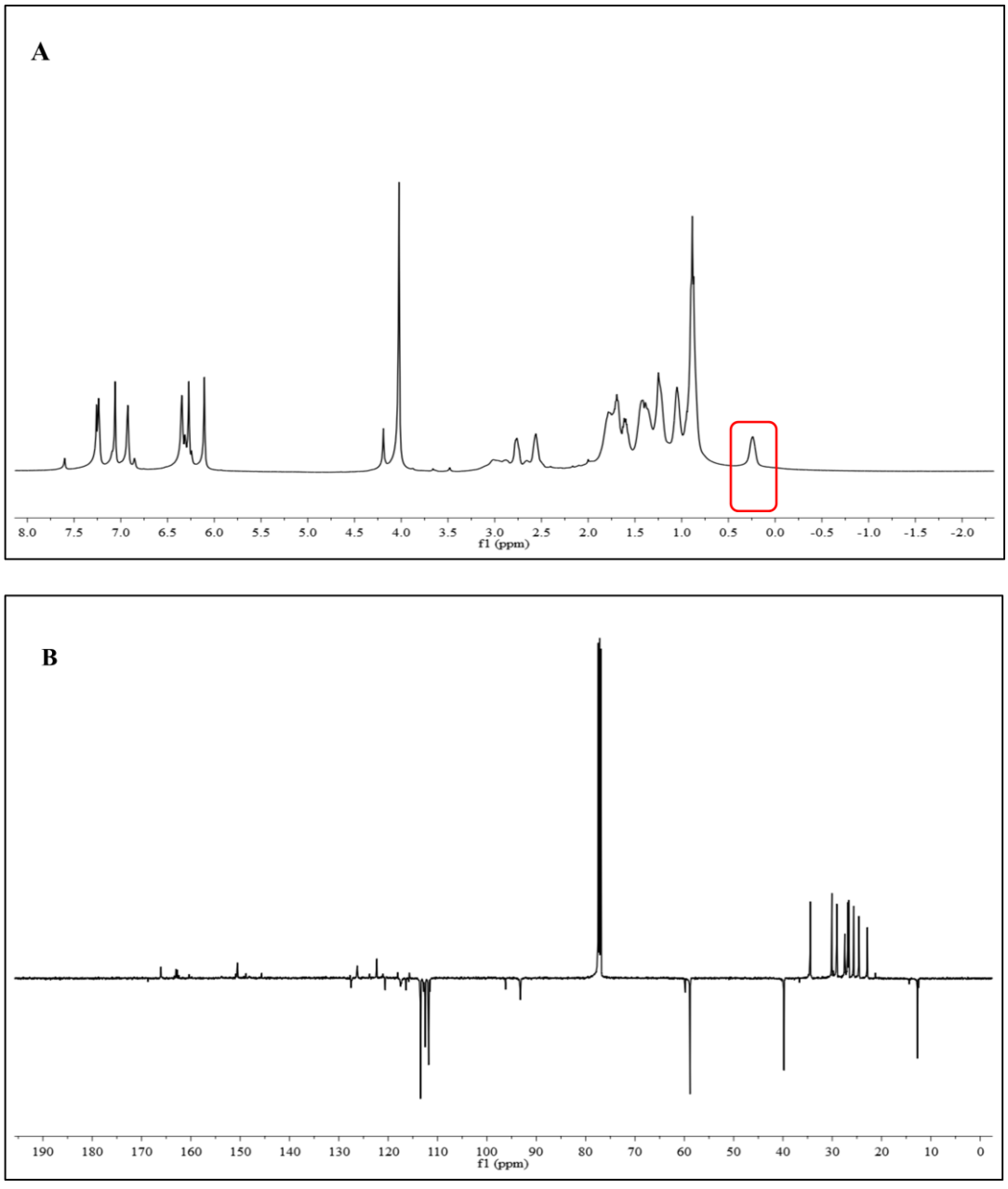


Figure 2.17 A. ^{1}H NMR spectrum in $CDCl_3$ (700 MHz) of metacycloprodigiosin. The characteristic signals at 0.2 ppm for metacycloprodigiosin (**19**) is highlighted by a box, a- methane protons, b- hydrocarbon protons α to the pyrrole ring. B. ^{13}C NMR spectrum ($CDCl_3$, 175 MHz) of metacycloprodigiosin (**19**).

These data confirmed the identity of the cyclic metabolite purified from *S. coelicolor* W31 expressing *mcpG* (from *S. longispororuber*) as metacycloprodigiosin (**19**) and the role of McpG in the cyclisation of undecylprodigiosin to form metacycloprodigiosin.

In conclusion, *mcpH* from *S. longispororuber* was expressed in *S. albus* and the condensation of synthetic 2-UP and MBC in the strain fed with was confirmed.

The previous construct to address the function of McpG was sequenced and found truncated. However, the truncated-McpG gave matacylprodigiosin in *S. albus* fed with undecylprodigiosin showing that the missing amino acids in the C-terminus of the protein could be not very essential for its activity.

The *mcpG* sequence was corrected and recloned. Heterologous expression of *mcpHG* and *mcpG* in *S. albus* and *S. coelicolor redG* mutant respectively gave cyclised metacycloprodigiosin (**19)** which was confirmed by NMR.

3. Investigation of *redG* orthologues from *Streptomyces griseoviridis*

3.1 *rph* gene cluster

As mentioned in chapter 1, prodigiosin R1 (**20**) and roseophilin (**21**) are produced by *Streptomyces griseoviridis*. Their biosynthesis is directed by the *rph* gene cluster which contains 21 genes homologous to genes in the *red* cluster of *S. coelicolor*. Within the cluster there are 4 *redG* orthologues (*rphG*, *rphG2*, *rphG3* and *rphG4*) (Figure 3.1). Examination of the carbon-carbon bond forming steps required to get to prodigiosin R1 (**20**) and roseophilin (**21**) from 11-methyldodecylprodigiosin (**26**) suggests that at least three RedG orthologues are required (Hayakawa *et al*., 2009). The sequences of RedG and the RphG proteins showed high similarity (Figure 3.2). Three *rphG*s (1, 2 and 4) contain the conserved [2Fe-2S] cluster binding motif characteristic of Rieske oxygenases (Figure 3.2).

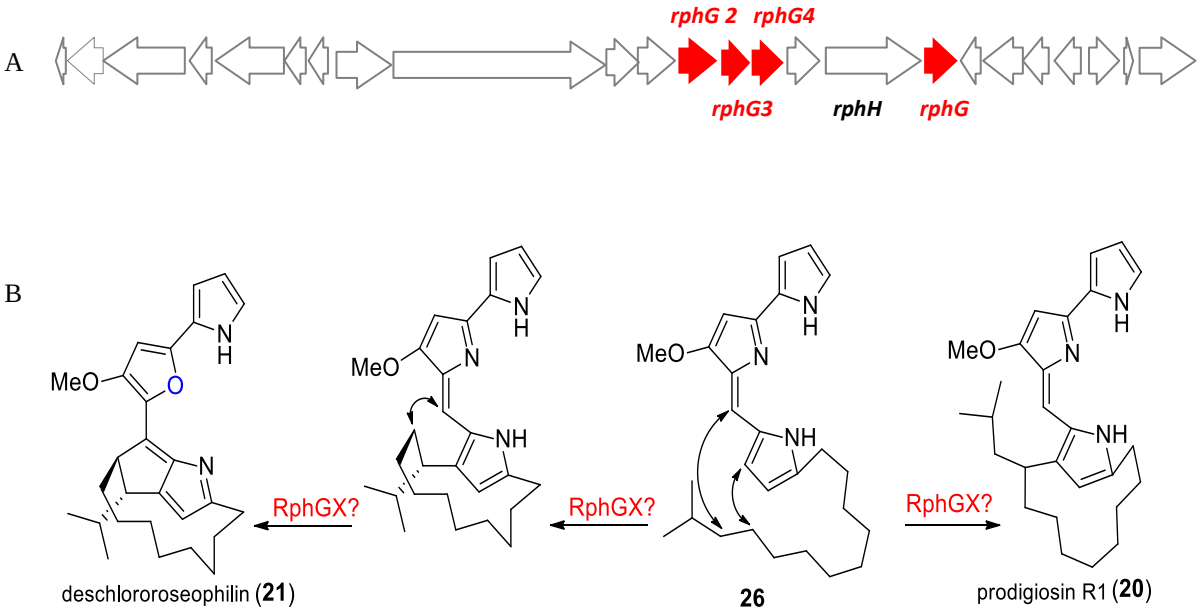


Figure 3.1 A. *rph* cluster proposed to encode prodiginine biosynthesis in *Streptomyces griseoviridis*. B. Three C-C bond forming steps required for the biosynthesis of roseophilin and prodigiosin R1 (**20**).

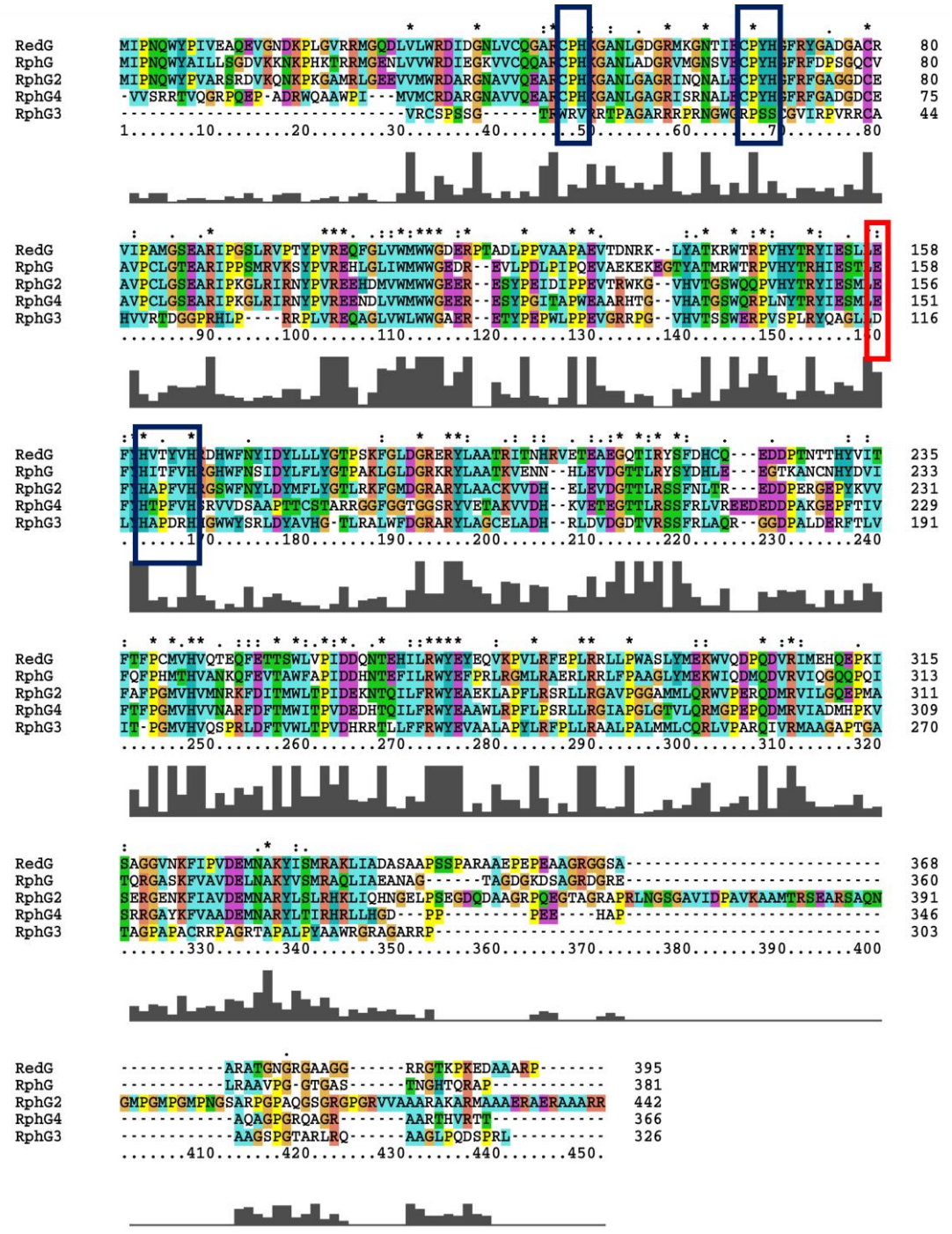


Figure 3.2 Protein sequence alignment of RedG and RphG, RphG2, RphG3 and RphG4. The conserved residues highlighted in blue box ligate the [2Fe-2S] cluster (missing in RphG3) and Fe(II) atom. A Glu residue in RedG and the RphGs, highlighted by a red box, is proposed to participate in electron transfer between the [2Fe-2S] cluster and the Fe(II) ion.

3.2 Heterologous expression of four *redG* orthologues; *rphG, rphG2, rphG3* and *rphG4* in *Streptomyces albus*

To investigate the integrity of the plasmids containing the *rphG* genes in pOSV556, which had previously been prepared by Paulina Sydor, *E. coli* TOP10 was separately transformed with each plasmid and the DNA was re-isolated. The integrity of each isolated plasmid was confirmed by restriction digestion with *Bam*HI (Figure 3.3) as well as by DNA sequence analysis (GATC).

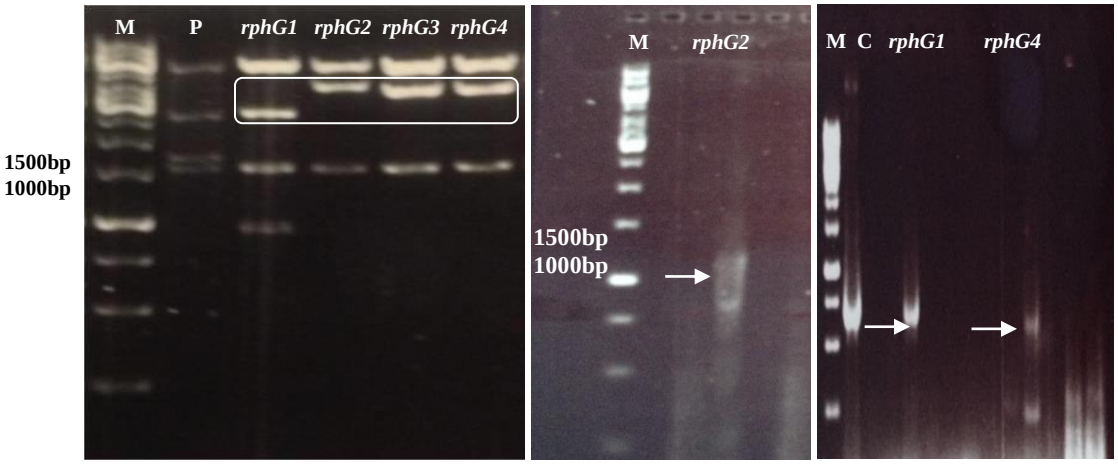


Figure 3.3 A. Agarose gel electrophoretic analysis of restriction digestion of pOSV556/*rphG1*, pOSV556/*rphG2*, pOSV556/*rphG3* and pOSV556/*rphG4* with *Bam*HI, 1146bp, 1329bp, 981bp, and 1101bp, respectively. The lane marked M contains 1kb molecular size. B and C. PCR amplification of *rphG1*, *rphG2*, and *rphG4* from *S. albus* transconjugants. The lane marked M contains1kb molecular size markers. PCR amplification of *rphG3*. The lane marked M contains1kb molecular size markers. P is pOSV556 and C is control.

Sequencing analysis showed that *rphG*3 appeared to be not correct. The construct pOSV556/*rphG*3 was therefore made again. The gene was amplified by PCR and the reaction was analysed by gel electrophoresis (Figure 3.4). The band with the correct molecular weight (981bp) was purified from the gel and digested with *Stu*I and *Pst*I. The digested PCR product was ligated with pOSV556 that had been digested with the

same enzymes and the digestion mixture was used to transform *E. coli* TOP10 cells selecting for ampicillin resistance.

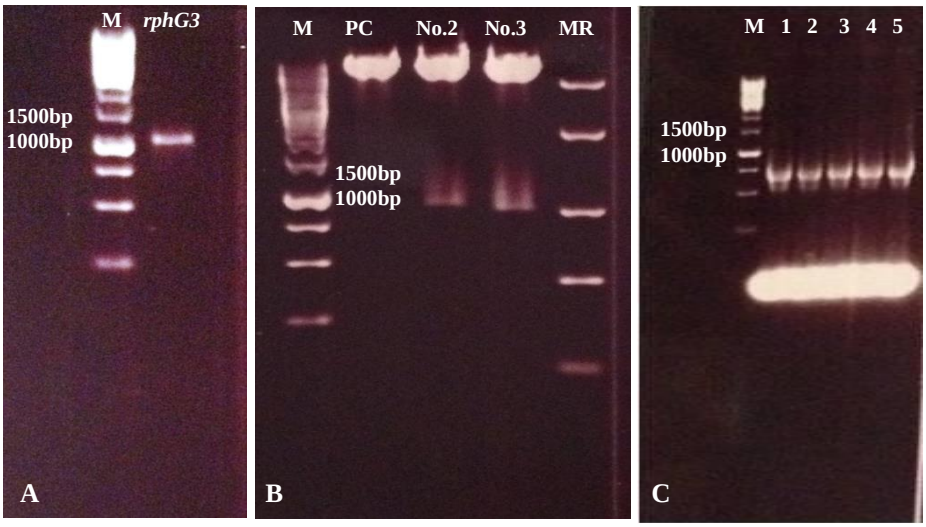


Figure 3.4 Agarose gel electrophoretic analysis of A. *rphG*3 PCR product, **B**. Restriction digestion of pOSV556 (vector) and pOSV556/*rphG*3 clone no.2 and clone no.3. **C.** *rphG*3 PCR from *S. albus* transconjugants (981bp). The lane marked M contains 1kb molecular size markers and the MR contains middle range molecular size markers.

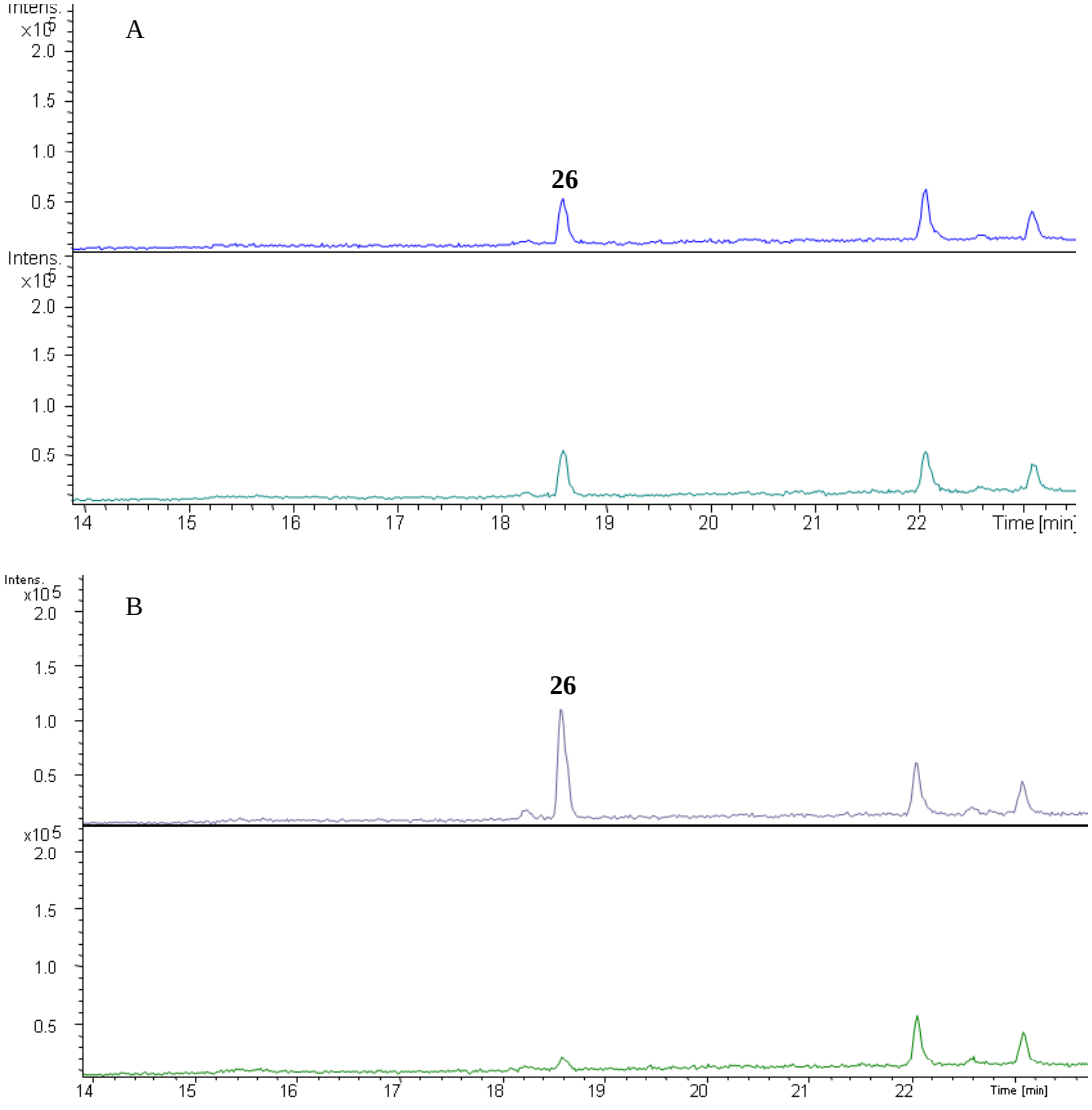
Plasmids isolated from clones were analysed by restriction digestion (Figure 3.4) and DNA sequence analysis. Unfortunately, a part of the gene was not covered by the sequencing results. This may be due to secondary structure at the 3' end of the gene. Since the rest of the sequence data showed correct insertion of *rphG3* the plasmid was used for subsequent experiments.

The restriction digestion analysis and DNA sequence data confirmed that all of the constructs apart from pOSV556/*rphG3* were correct. In order to express all four plasmids in *S. albus* (as this was confirmed to be a superior host than *S. coelicolor* (Figure 2.5, Section 2.4), *E. coli* ET12567/pUZ8002 was separately transformed with each plasmid. Ampicillin resistant colonies were picked and conjugation was performed between *E. coli* transformants and *S. albus*, selecting for hygromycin

resistance. Colonies were picked and analysed by PCR for the presence of the appropriate gene (Figure 3.3 and Figure 3.4).

3.2.1 Feeding with synthetic 11-methyl-dodecylprodigiosin

These transconjugant *Streptomyces* strains expressing *rphG*, *rphG2*, *rphG3* and *rphG4* were grown on R5 medium on a semi-permeable membrane and fed with 11-methyldodecylprodigiosin (synthesised by David Withall, another PhD student in the group). The extracts were analysed by LC-MS (Figure 3.5).



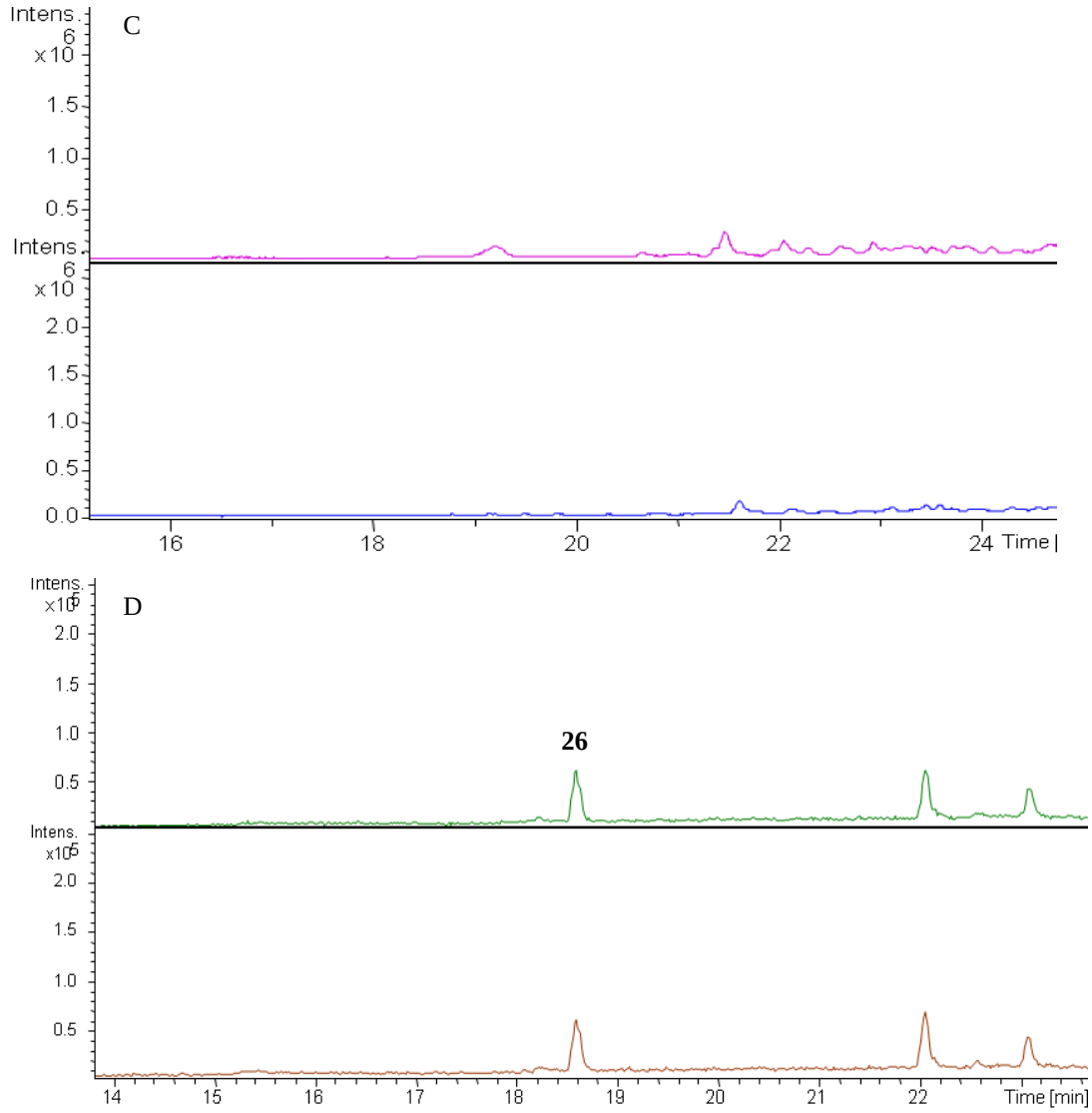


Figure 3.5 Activity of RphGs in *S. albus*. EIC for *m/z* range 392-422 (corresponding to $[M+H]^+$ for 11-methyl-dodecylprodigiosin (*m/z* = 422.31) from LC-MS analysis of acidified organic extract of A. *S. albus*. B. *S. albus*/pOSV556/*rphG*1. B. *S. albus*/pOSV556/*rphG2*. C. *S. albus*/pOSV556/*rphG*3. D. *S. albus*/ pOSV556/*rphG*4 (top traces of A, B, C and D are *S. albus*) fed with synthetic 11-methyl-dodecylprodigiosin (*m/z* = 422.3039).

Unfortunately, no carbocyclic derivatives of 11-methyldodecylprodigiosin (**26**) could be observed in any of the extracts. To examine whether the host strain is the reason for lack of activity, we planned to express them in the *S. coelicolor redG* mutant. Although RedG utilises a substrate with a different alkyl chain it was hoped this would still act as a substrate for one or more of the RphG enzymes.

3.3 Heterologous expression of *redG* orthologues: *rphGs* in *S. coelicolor* W31 (M511Δ*redG*::*scar*)

The four vectors were separately transferred into *S. coelicolor* W31, which produces undecylprodigiosin (**2**) but no streptorubin B (**3**) (Figure 3.6).

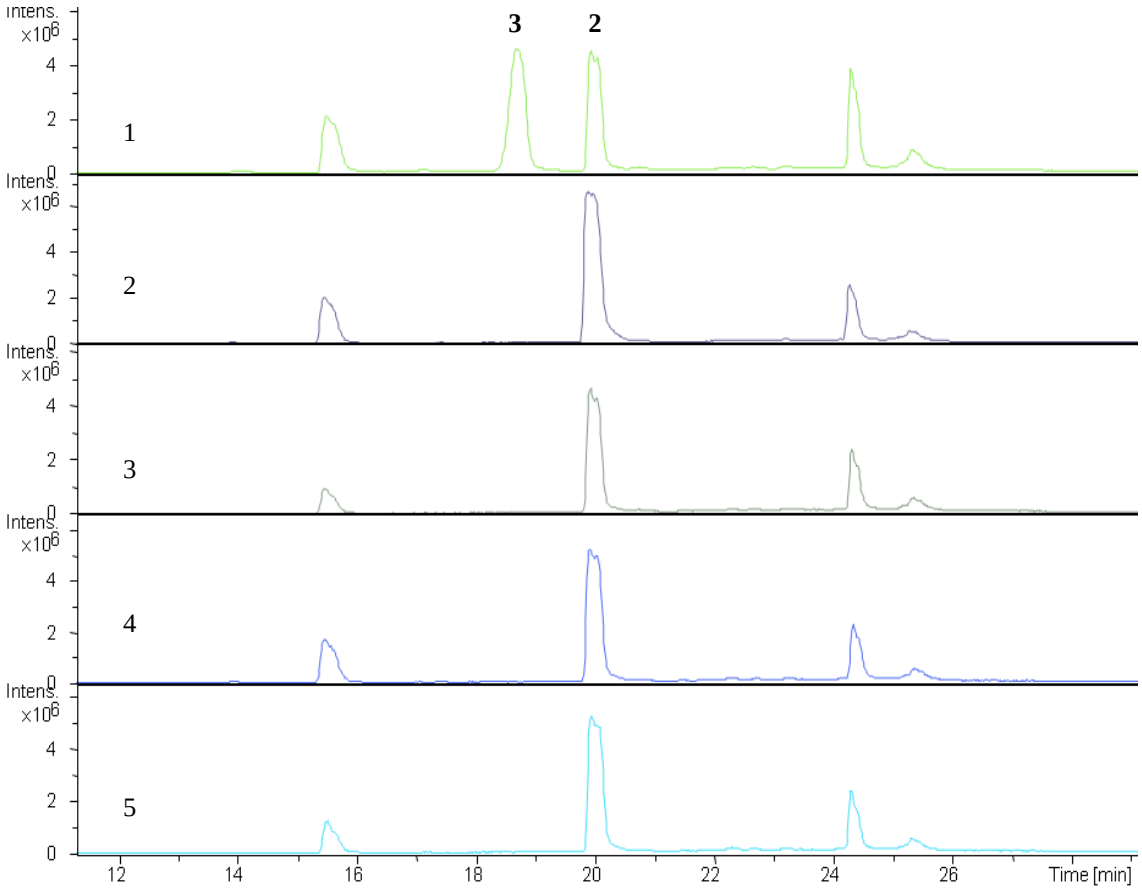


Figure 3.6 Activity of RphGs in *S. coelicolor* W31Δ*redG*. A. EIC for *m/z* = 392-394 (corresponding to $[M+H]^+$ for undecylprodigiosin (**2**) and metacycloprodigiosin (**19**)/streptorubin B (**3**) from LC-MS analysis of acidified organic extracts of 1. *S. coelicolor* M511 wild type, 2. *S. coelicolor* W31 (Δ*redG*), 3. *S. coelicolor* W31 /pOSV556*rphG*1, 4. *S. coelicolor* W31/pOSV556*rphG*2 and 5. *S. coelicolor* W31 /pOSV556*rphG*4.

Unfortunately, no oxidative carbocyclisation products were observed when *rphG*s were expressed alone. The aim of this experiment was to see whether the RphGs can catalyse oxidative carbocyclisation reactions of **2**. The RphGs were not functional in *S. coelicolor* W31 (Figure 3.6) possibly because **2** is not the natural substrate.

3.4 Cnstruction of expression vectors for *rphH*, *rphH/rphG*, *rphH/rphG2*, *rphH/rphG3* and *rphH/rphG4*

As previously described there is some evidence to suggest that RedG forms a complex with RedH in *S. coelicolor* that enhances its catalytic activity. Thus an alternative approach aimed at investigating whether RphH boosts the activity of the RphG enzymes was pursued. To investigate this possibility, *rphH* (3252bp) was synthesised by Epoch Life Science with a synthetic ribosomal binding site (RBS) AGGAGG and 7bp before the start codon. *Pml*I and *Hin*dIII restriction enzymes were added to the 5' and 3' ends of the construct, respectively (Figure 3.7).

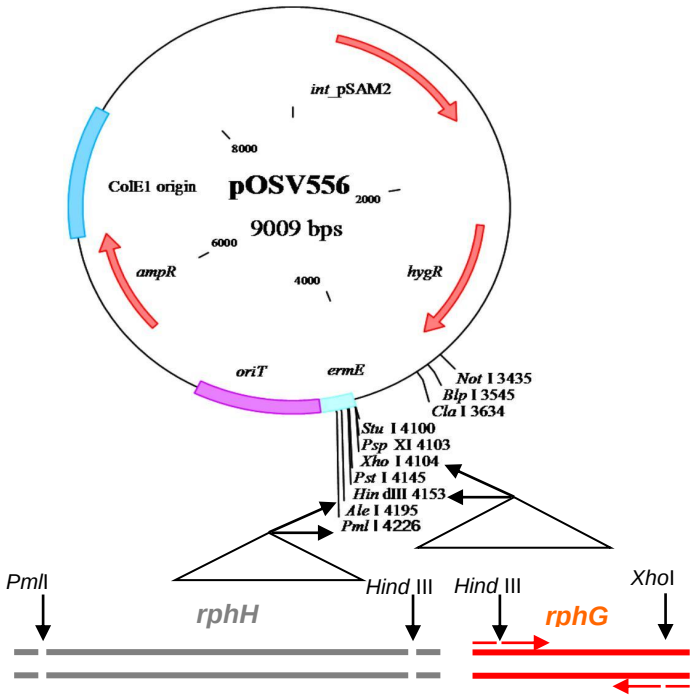


Figure 3.7 Cloning of synthetic *rphH* gene into previously prepared pOSV556 containing *rphG1*, *rphG*2, *rphG*3 or *rphG*4 separately.

The gene was amplified by PCR, cloned into pOSV556 and confirmed by digestion and DNA sequence analysis. *rphH* was then sub-cloned into pOSV556/*rphG*, pOSV556/*rphG*2, pOSV556/*rphG*3 and pOSV556/*rphG*4 (Figure 3.7).

The cloning was confirmed by restriction digestion and sequencing (Figure 3.8). *S. albus/*pOSV556/*rphHG*, *S. albus/*pOSV556/*rphHG*2, *S. albus/*pOSV556/*rphHG*3, *S. albus/*pOSV556/*rphHG*4 were made using standard procedures.

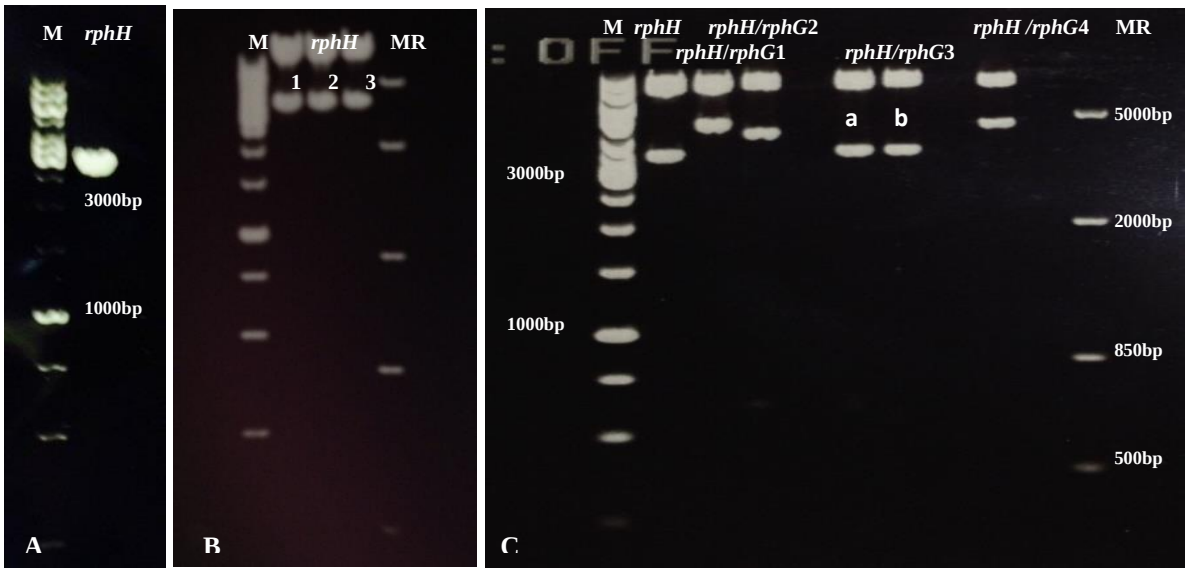


Figure 3.8 Agarose gel electrophoretic analysis, A. PCR product of *rphH*. B. Digestion of synthetic *rphH* cloned into pOSV556 with *Pml*I and *Hin*dIII. C. Digestion of sub-cloned synthetic *rphH* gene into previously recombinant pOSV556 holding *rphG*1, *rphG* 2, *rphG*3 and *rphG*4 separately, with *Pml*I and *Xho*I. The lane marked M contains 1kb molecular markers and the MR contains middle range molecular markers.

3.4.1 Feeding of synthetic 11-methyldodecylpyrrole and MBC to *S. albus* expressing *rphH*

Before conducting feeding experiments to determine whether RphGs require RphH for activity, the activity of RphH alone was determined. The plasmid *rphH*/pOSV556 was introduced into *S. albus* by conjugation as previously described and 11-

methyldodecylpyrrole (**25**), synthesised by Douglas Roberts (a research fellow in Challis group), and MBC (**16**) were fed to the resulting strain. Unfortunately, no condensation of **25** and **16** was observed. A positive control using McpH from *S. longispororuber* afforded a condensation product even when **25** was used, instead of 2-UP (**15**), with **16** (Figure 3.9).

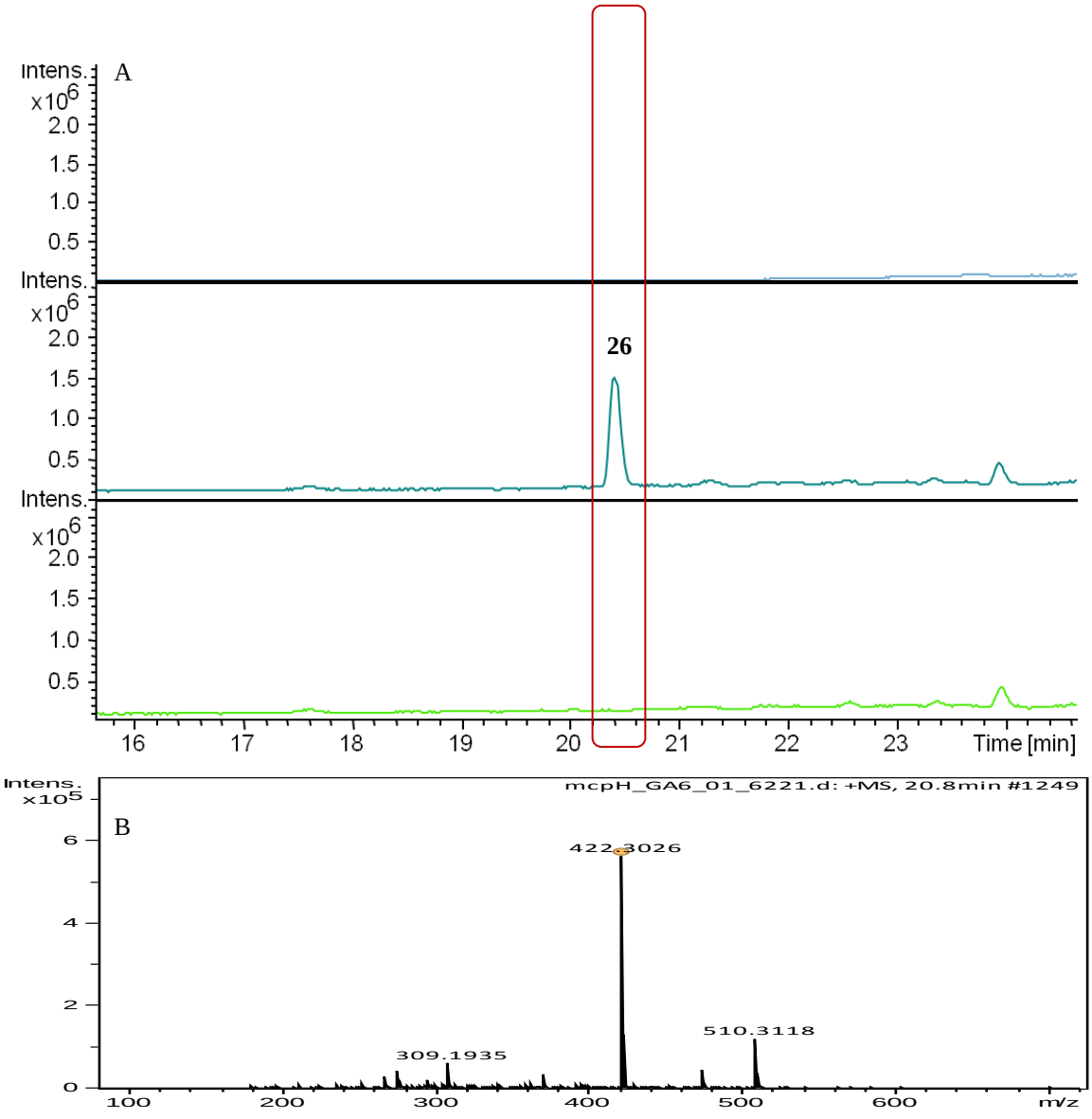


Figure 3.9 Activity of RphH from *S. griseoviridis*. A. EIC for *m/z* = 422 (corresponding to $[M+H]^+$ for 11-methyldodecylprodigiosin (**26**) from LC-MS analysis of acidified organic extract of *S. albus* (top trace), *S. albus*/pOSV556/*mcpH* (middle trace) and *S. albus*/pOSV556/*rphH* (bottom trace) fed with 11-methyldodecylpyrrole and MBC. B. HR-MS confirming the molecular formula of **26** (calculated for $C_{27}H_{39}N_3O$= 422.3039, found; 422.3026).

3.4.2 Cloning and expression of *rphH* variants in *S. albus*

To investigate the potential reason for the inactivity of RphH, a protein sequence alignment for RphH was performed against RedH from *S. coelicolor* and McpH from *S. longispororuber*. The 139 amino acids of the N-terminus of RphH showed no correlation to RedH or McpH (Figure 3.10A). Furthermore, two alternative start codons can be observed at the start of the region of homology (Figure 3.10B).

A

.....*

B

GTGTCTGTGCCGGCGGCGCCGGACGGGTCCGCCGCTCCGCCTGCGGAGGGGACGGACGCTTCCGCCAGGTC
CGCGGCGGAGCCGCGGGGTTCTTCGGCGGCGGCGGCGGCGCGGGGCGCCGCGGGTTCCGGGGGTGACCCC
GTGGGACCGGTCGCGCGGAACGCGACGGCCTCGTCCGCGGTGGAGCCGCGGGAGGTGGCCGGGGCCGGG
GATGATGCCGCCCGTCAGGGCGGGAAGGCGGCACGGTCGCGGGCGAAGAAGGCGGCGGGCGGTGACCCGT
CGCGCAAGGCACCCGCCGCGAAGGCGGCCAGGCGGAAGACGGCCGCACCGTCCCGCCGTTCCCCCGCCGGC
ACCAGCCCGGCGGCGGAGCCGCCGGAGCCCGCCGCCGCTGCGGCGGGAACGCGCGCCCCCGCGCCGCCCGC
GGATGGCCATGGTCCGCCCACCGGGCCCGGCACGGCGGCGGAGCCGCCCGCCCGCGCCCGGGCCGGCGCGT
CCGCCGCGGCGCCGGAGGCGCCGTCGGCGGACGTCGCCGTGCTCGGCGCCGACGCCGCGGCGAAGCCGGA
AACGCTGGGGGGCAAGGCCGCCCGTCTCGGCGAGATGATCGAAGCCGGCTTCCCCGTGCCGCCCGCCGTCTG
CCTGACCACCGAGCTGTTCCACCGCTTCCTGCGCGAGACCGGCCTGGACAAGG.....
.....
.....TGA

Figure 3.10 A) Protein sequence alignment of the RedH, McpH and RphH, The two places where the different RphH variants begin are marked in red. B) DNA sequence of *rphH*, the two alternative start codons are marked in green.

It is possible that this extra protein sequence is not necessary for catalysis and it impairs the activity. Therefore, two RphH variants were made using two different start codons (highlighted in Figure 3.10) which gave two possible constructs that cover the conserved amino acids in the N-terminus of the three proteins. The RphH variants were made using standard cloning procedures and the integrity of each mutant was confirmed by digestion (Figure 3.11) and sequencing.

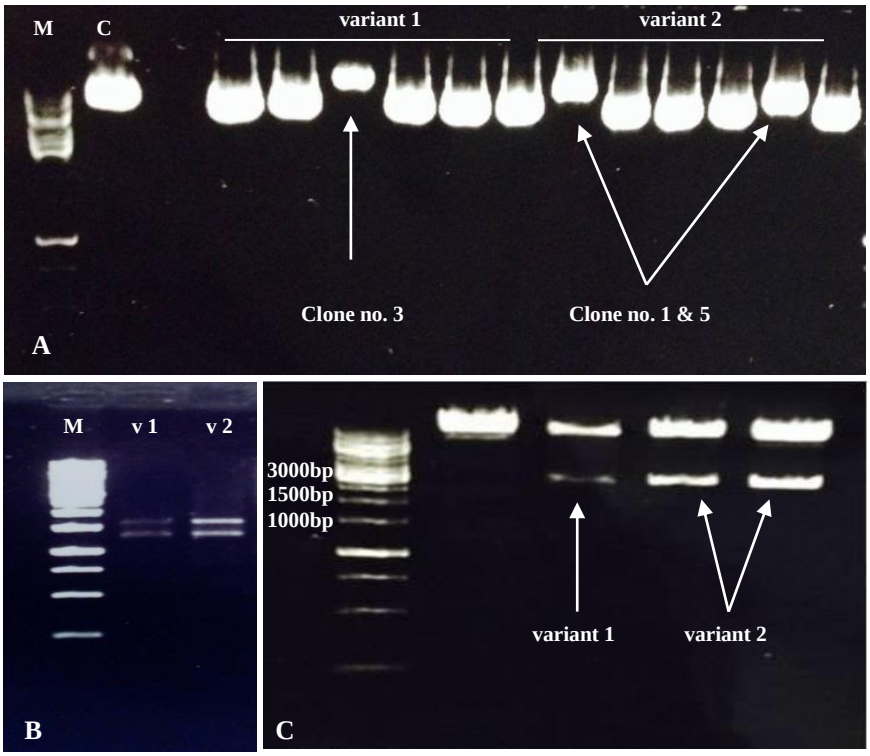


Figure 3.11 Agarose gel electrophoretic analysis, A. Single digestion of 12 clones of pOSV556/*rphH* variant 1 and pOSV556/*rphH* mutant2. B. PCR product of *rphH* variant 1 and variant 2 digested with *Kpn*I. C. pOSV556/*rphH* variant 1 and pOSV556/*rphH* variant 2 digested with *Pml*I and *Hin*dIII.

3.4.3 Feeding of synthetic 11-methyldodecylpyrrole and MBC to *S. albus* expressing *rphH* variants

Following the conjugation into the *S. albus* strains expressing *rphH*, *rphH*-mutant 1, *rphH*-mutant 2 and *mcpH* from *S. longispororuber* (as a control) were fed with

synthesised 11-methyldodecylpyrrole (**25**) and MBC (**16**). LC-MS analysis showed that no peaks corresponding to a condensation product could be observed for any of the RphH variants (Figure 3.12). The strategy of expressing *rphHG*, *rphHG2*, *rphHG3* and *rphHG4* was therefore not pursued.

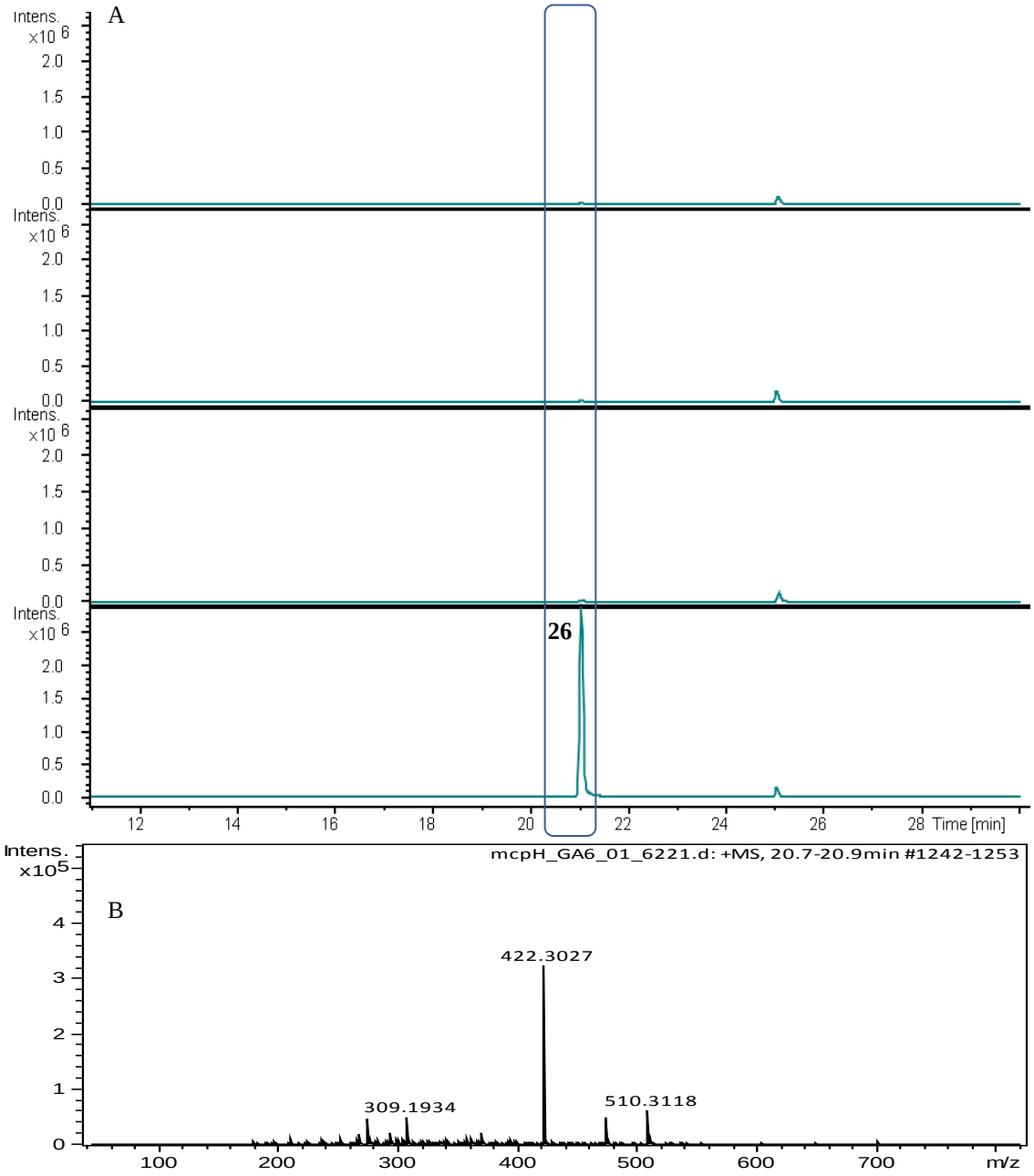


Figure 3.12 Activity of RphH from *S. griseoviridis*. A. EIC for *m/z* = 422 (corresponding to $[M+H]^+$ for (**26**) from LC-MS analysis of acidified organic extract of *S. albus*/pOSV556/*rphH* (top), *S. albus*/pOSV556/*rphH* variant 1 (second from top) *S. albus*/pOSV556/*rphH* variant 2 (second from bottom) and *S. albus*/pOSV556/*mcpH* (bottom) fed with 11-methyldodecylpyrrole and MBC. B. HR-MS confirming the molecular formula of **26** (calculated for $C_{27}H_{39}N_3O$= 422.3039, found; 422.3027).

3.5 Expressing *rphG* in *Streptomyces coelicolor* W31 (M511Δ*redG::scar*) using a universal RBS

In an attempt to improve the expression level of *rphG*, a new construct containing a universal ribosomal binding site (RBS) (AGGAGG) was made and expressed in *S. coelicolor* W31.

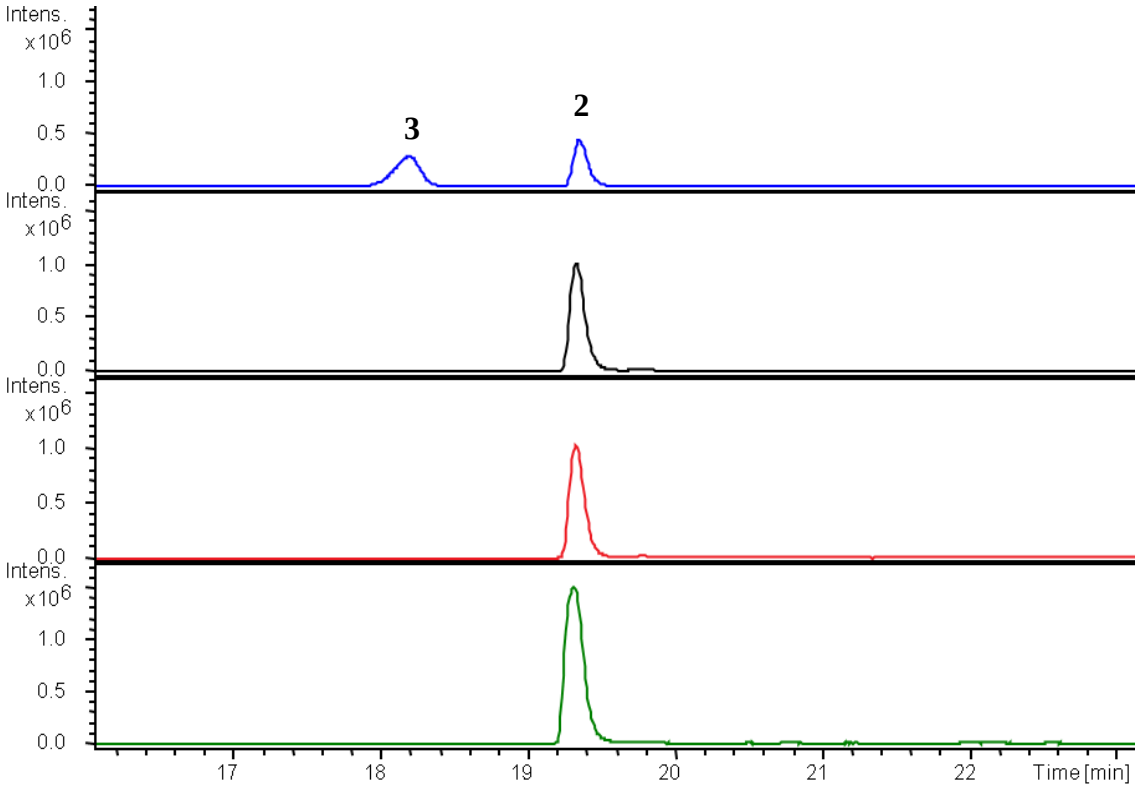


Figure 3.13 Activity of RphG from *S. griseoviridis*. A. EIC for *m/z* =422 (corresponding to $[M+H]^+$ for undecylprodigiosin (**2**) and metacycloprodigiosin (**19**)/streptorubin B (**3**) from LC-MS analysis of acidified organic extract of *S. coelicolor* M511 wild type (top), *S. coelicolor* Δ*redG* (second from top), *S. coelicolor* Δ*redG* pOSV556/*rphG* with natural RBS (second from bottom) and *S. coelicolor* Δ*redG* pOSV556/*rphG* with universal RBS (bottom).

Unfortunately, this variant did not produce any carbocyclised undecylprodigiosin derivatives (Figure 3.13).

3.6 Re-cloning corrected *rphG* and expressing in *Streptomyces coelicolor* W31 (M511Δ*redG::scar*)

The *rphG* sequence was recently corrected by the Kawasaki group, who showed that the corresponding protein contains an additonal 59 amino acids at its C-terminus (Figure 3.14) (Hayakawa *et al*., 2009; Hayakawa *et al*., 2016). This corrected sequence was assembled from a series of overlapping synthetic oligonucleotides by Chuan Huang (a research fellow in the Challis group) using overlapping oligos.

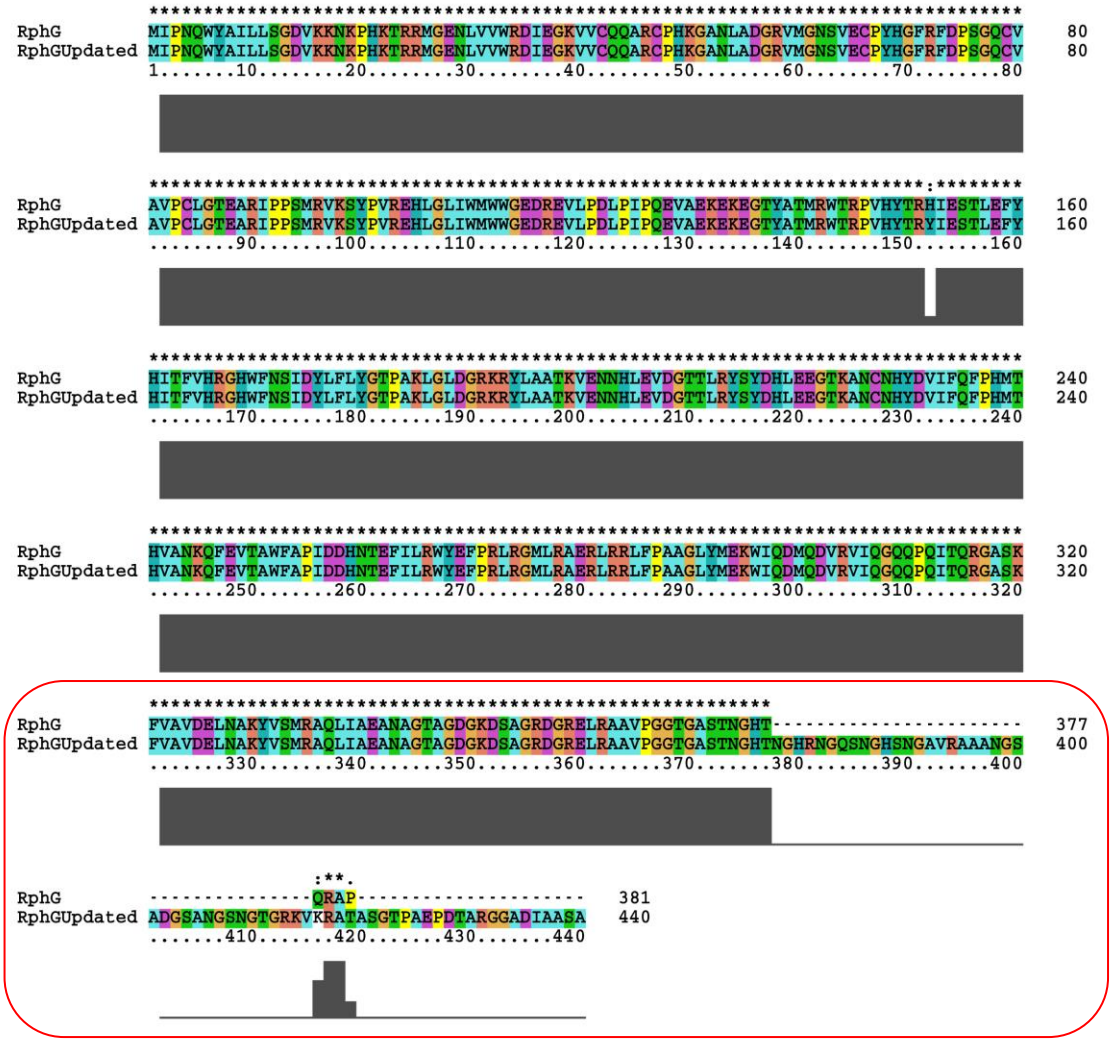


Figure 3.14 Alignment of the original (2009) and corrected (2016) RphG sequences showing 59 additonal amino acids are appended to the C-terminus of the protein.

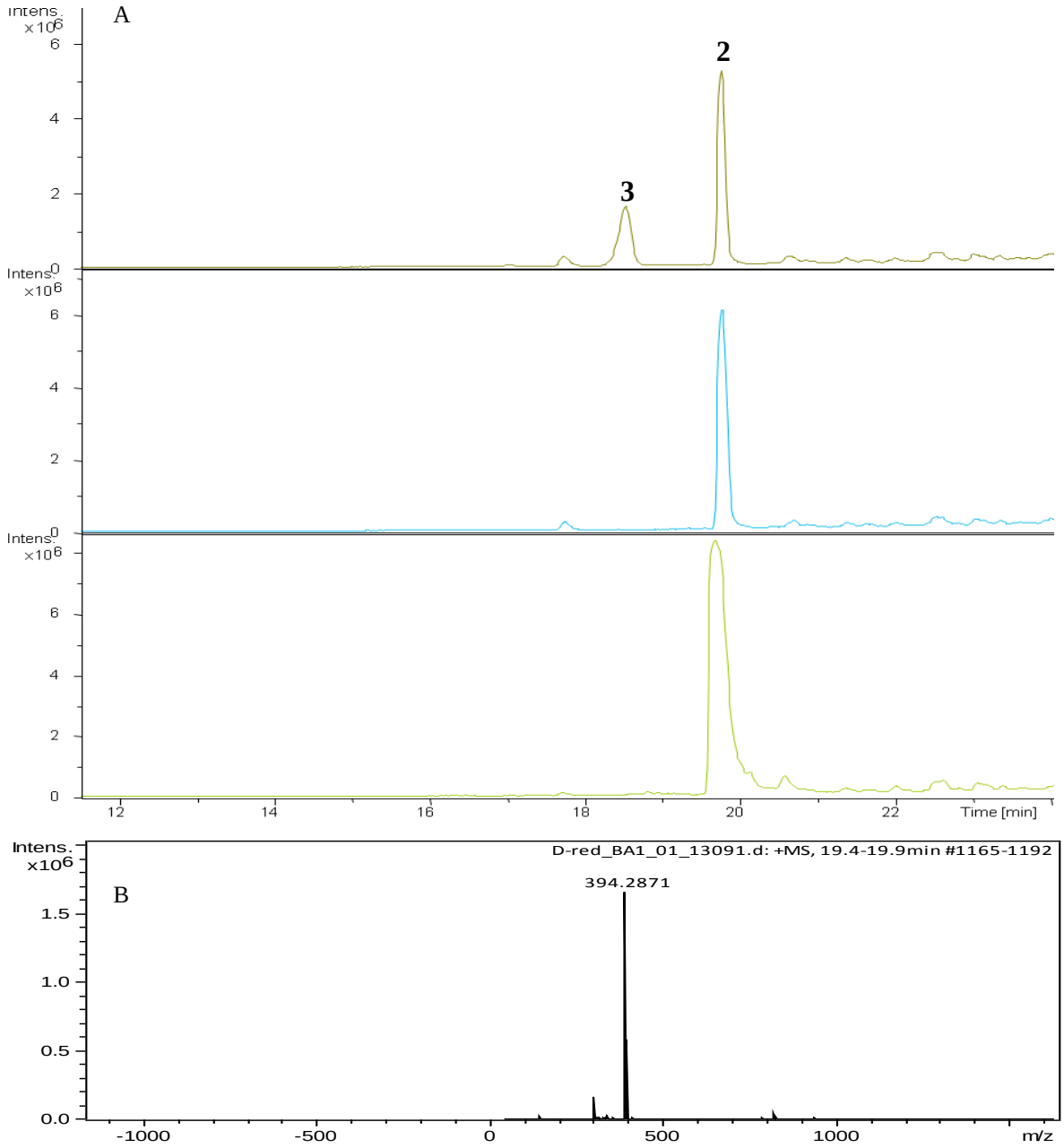


Figure 3.15 Activity of RphG from *S. griseoviridis*. A. EIC for *m/z* =422 (corresponding to $[M+H]^+$ for undecylprodigiosin (**2**) and metacycloprodigiosin (**19**)/streptorubin B (**3**) from LC-MS analysis of acidified organic extract of of *S. coelicolor* M511 wild type (top trace), *S. coelicolor* Δ*redG* (middle trace) and Δ*redG* pOSV556/*rphG* (bottom trace). B. HR-MS confirming the molecular formula of **2** (calculated for $C_{25}H_{37}N_3O$= 394.2856, found; 394.2871).

The *rphM* promoter sequence from *S. griseoviridis* was used for the expression. Neither *rphG2*, *rphG3* nor *rphG4* gave any cyclised metabolite. The correct *rphG* sequence was expressed in a *redG* mutant of *S. coelicolor* M511. Unfortunately, even expression of this corrected *rphG* sequence did not result in the production of any carbocycyclic undecylprodigiosin derivatives (Figure 3.15), possibly because a different promoter (*ermE*) from the one used by the Hayakawa group was used to express the gene.

Meanwhile, Hayakawa and co-workers (Kimata *et al*., 2016) expressed the correct *rphG* sequence in a *redG* mutant of *S. coelicolor* M511 which resulted in the production of two carbocylic undecylprodigiosin derivatives (metacycloprodigiosin and propyl-meta-cyclooctylprodiginine).

In conclusion, expressing *rphGs* in *S albus* and *S. coelicolor redG* mutant did not result in the production of any carbocyclic prodiginines. RphH and its variants were not able to catalyse the condensation of MBC with 2-UP or deriatives whereas *mcpH* (used as a control) from *S. longispororuber* could catalyse such reactions. The sequence of *rphG* was corrected by the Hayakawa and co-workers (2016) while our work was in progress, suggesting that there may also be problems with the sequences of *rphH* and the other three *rphG* genes.

Whole genome sequence of *Streptomyces longispororuber*

4. Whole genome sequencing of *Streptomyces longispororuber*: Insight into specialised metabolites biosynthesis

As discussed in chapter 2, the amount of cyclised undecylprodigiosin formed following expression of *mcpG* from *S. longispororuber* in a heterologous host was very low compared to the analogous experiment with *redG* from *Streptomyces coelicolor*. Sequencing of *mcpG* amplified from a fosmid (from lab collection) showed that the predicted stop codon did not appear to be a stop codon. In addition, protein sequence comparison of McpG with RedG from *S. coelicolor* showed that McpG was 48 a. a. shorter than the RedG. In order to be completely certain of the McpG sequence and to have an insight into other putative natural product biosynthetic gene clusters, the whole genome of the *S. longispororuber* strain was sequenced.

The DNA sample was prepared using a standard DNA extraction protocol (chapter 6) and was sequenced by the Erlham Institute.

4.1 Metacycloprodigiosin biosynthetic gene cluster

The metacycloprodigioisn biosynthetic gene cluster appears to contain 23 genes (3332160 - 3400658nt). Compared to the *red* cluster from *S. coelicolor,* some genes from the metacycloprodigiosin cluster are located on the opposite DNA strand. All the genes observed in the prodiginine biosynthetic cluster in *S. coelicolor* are present within the McpG cluster, with the exception that the *redS* homologue (*mcpS*), which is not required for prodiginine biosynthesis in *S. coelicolor,* appears to not be present in *S. longispororuber*. Additionally, the homologues of *redL* and *redK* appear to be fused together to form one gene *mcpKL* (Figure 4.1).

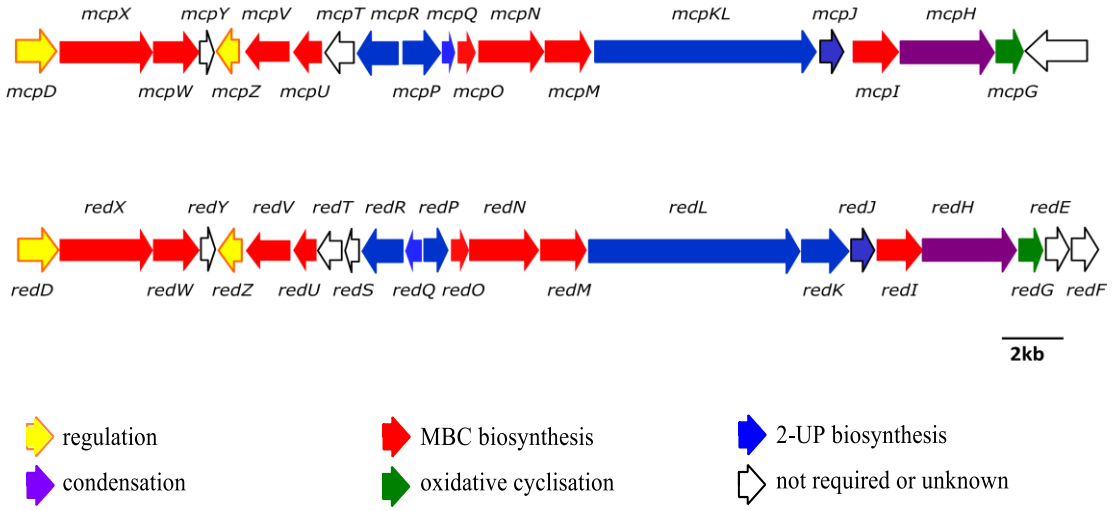


Figure 4.1 Cluster cluster alignment; *mcp* cluster from *S. longispororuber* (upper) and *red* cluster from *S. coelicolor* (bottom).

Protein sequence comparision between McpH from fosmid 3G3 (Lab collection) and McpH from the sequenced strain (Figure 4.2) showed a few amino acid substitutions in McpH. These are due to errors in the originally determined sequence of *mcpH*.

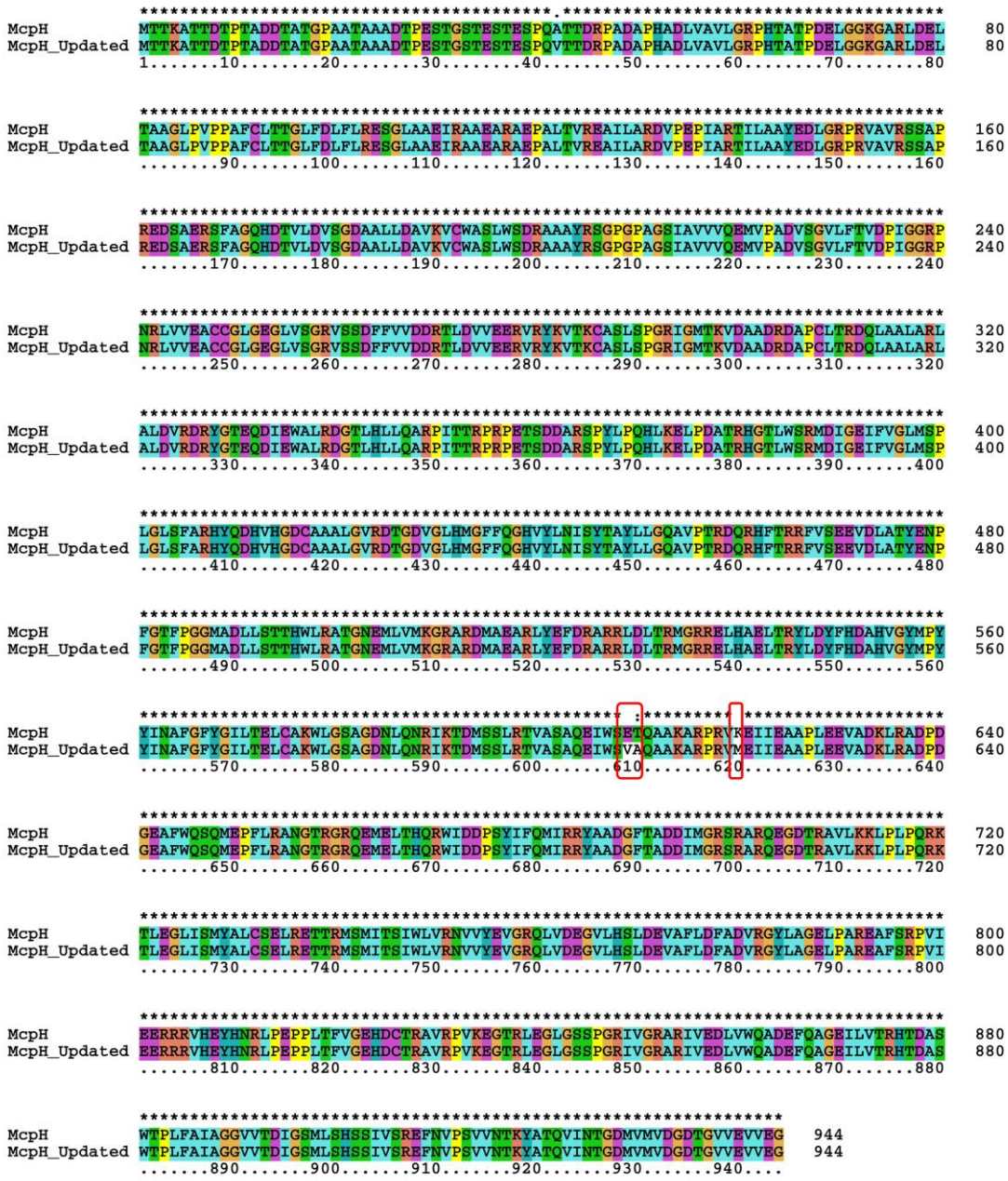


Figure 4.2 Protein sequence alignment of the McpH and McpH (updated) from *S. longispororuber*.

Protein sequence comparison between McpG from fosmid 3G3 (Lab collection) and McpG from the sequenced strain (Figure 4.3) showed that 66 amino acids were missing from the clones amplified from the fosmid. This confirmed the original protein used in the heterologous expression of *mcpG* was truncated and allowed for preparation of a construct with the correct sequence (chapter 2).

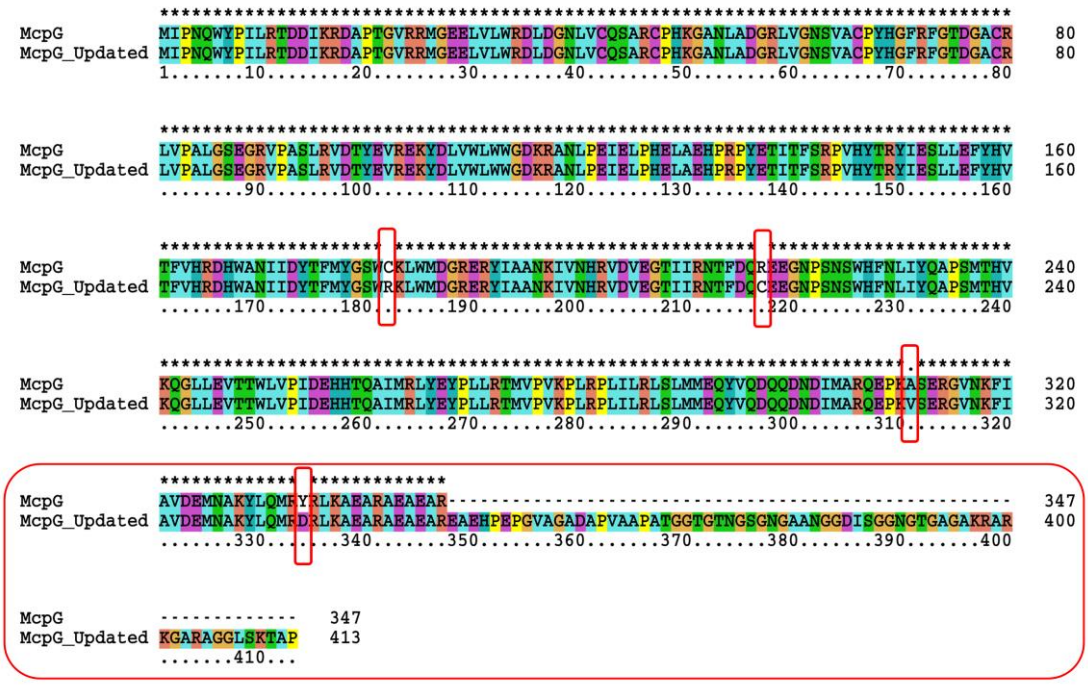


Figure 4.3 Protein sequence alignment of the McpG and McpG (updated) from *S. longispororuber*.

4.2 Analysis of the biosynthetic potential of *S. longispororuber*

The only natural product currently known to be produced by *S. longispororuber* is metacyloprodigiosin. Since most *Streptomyces* strains produce several specialised metabolites, the genome sequence of *S. longispororuber* was analysed to illuminate the biosynthetic potential of this strain. In collaboration with Dr Emmanuel Lorenzo de los Santos antiSMASH (Weber *et al*., 2015) was used to identify biosynthetic loci, and the sequence (9,239,768 bp) revealed many cryptic gene clusters for natural product biosynthesis. The closest homologue of each gene was then identified by BLAST analysis (appendix). From this analysis we were able to predict more than 30 gene clusters predicted to be responsible for the production of different natural products including six polyketides, five non-ribosomal peptides, and three hybrid PKS/NRPS products (Table 4.1 and Figure 4.4).

Table 4.1 Overview of gene clusters from *S. longispororuber* predicted by antiSMASH.

Cluster No.	Type	Metabolic product of most similar known cluster	Gene cluster Similarity[a] (%)
1	Melanin	Istamycin	5
2	Otherks	Thioviridamide	15
3	T2pks	Kinamycin	22
4	Ladderane-Nrps	Skyllamycin	14
5	Terpene	Albaflavenone	100
6	Phosphonate	-	
7	Siderophore	-	
8	Nrps	Mirubactin	78
9	T1pks	Hygrocin	22
10	Bacteriocin	-	
11	Lassopeptide	Asukamycin	3
12	Terpene	-	
13	Terpene	Bacterioaminohopanetriol	84 (100)
14	T1pks	Marineosin	81
15	T1pks	ECO-02301	53
16	Lassopeptide	Herboxidiene	3
17	Bacteriocin	-	
18	Lassopeptide	Chaxapeptin	42
19	Other	Salinomycin	4
20	T2pks	Pamamycin	100
21	Lantipeptide	-	
22	Lantipeptide-T1pks-	Polyoxypeptin	67
23	T1pks-Nrps	Pristinamycin	12
24	Nrps-Transatpks-	Kalimantacin /batumin	13
25	Nrps	Quartromicin	5
26	Other	Echosides	11
27	Terpene	2-methylisoborneol	100
28	Nrps	Nanchangmycin cluser	6
29	Bacteriocin-T1pks-	Telomycin	11
30	Nrps	Coelichelin	100
31	T3pks	Akaeolide biosynthetic gene cluster	12
32	Ectoine	Ectoine	100
33	Nrps	unknown	57
34	Nucleoside	Polyoxin biosynthetic gene cluster	5
35	Siderophore	Desferrioxamine	83 (100)
36	Melanin	Melanin biosynthetic gene cluster	28
37	Other	Staurosporine biosynthetic gene	35

[a] number in brackets are similarity scores resulting from manual inspection of the gene clusters

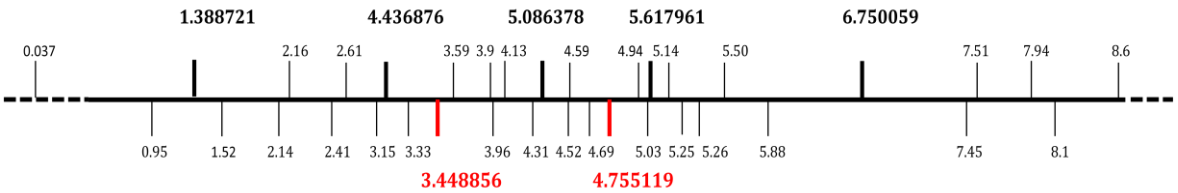


Figure 4.4 Location of gene clusters on the chromosome of *S. longispororuber* as predicted by antiSMASH; five clusters that showed 100% similarity (in black-bold) and two clusters appeared to be interesting; cluster 15 at 3.448856 nt and cluster 24 at 4.755119 nt (in red).

4.3 Gene clusters identical to known clusters from other microorganisms

Within the clusters, four clusters showed 100% similarity to known clusters from other *Streptomyces* strains, albaflavenone (cluster 5), pamamycin (cluster 20), 2-methylisoborneol (cluster 27) and ectoine (cluster 32). Additionaly, although gene cluster 13 and gene cluster 35 showed 84% and 83% similarity respectively, a closer inspection of the genes showed that they have 100% similarity to bacterioaminohopanetriol and desferrioxamine (Table 4.1 and Figure 4.5). With the exception of pamamycin, all of these clusters are also present in *S. coelicolor*; and in fact are found in most *Streptomyces* species (Challis, 2014). Albaflavenone, coelichelin and desferrioxamine have already been discussed in sections 1.24 and 1.25 resepectively. 2-methyl-isoborneol is produced by *Micromonospora olivasterospora* and is one of the secondary metabolites that are produced by *Streptomyces* and some other microorganisms that are responsible for the characteristic odour which arises from moist soil (Komatsu *et al*., 2007; Jiang *et al*., 2006). Ectoine which is produced by α-, γ-Proteobacteria and Actinobacteridae (Zhu *et al*., 2014). It is well known to serve as an osmoprotectant against increased external osmotic pressure by stabilising enzymes in the cell.

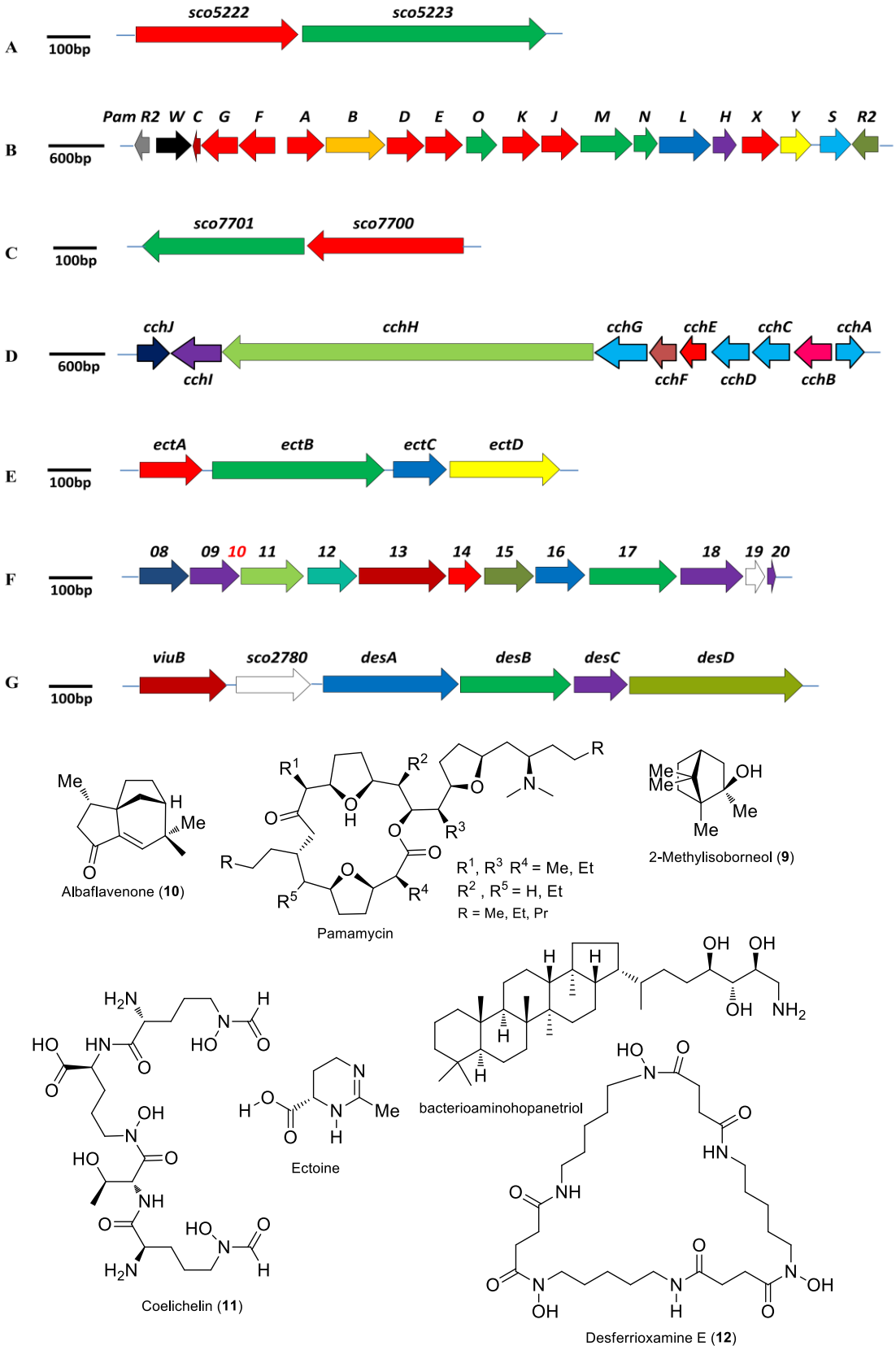


Figure 4.5 (Upper): Organisation of gene clusters from *S. longispororuber* that showed %100 similarity; A. Gene cluster 5, albaflavenone. B. Gene cluster 20, pamamycin. C. Gene cluster 27, 2-methyl-isoborneol. D. Gene cluster 30, coelichelin. E. Gene cluster 32, ectoine. F. Gene cluster 13, bacterioaminohopanetriol (gene 10 which is of unknown function is absent). G. Gene cluster 35, desferrioxamine. (Bottom): Structure of the metabolites encoded by the gene clusters.

Therefore, it has been suggested to be potentially useful commercially as an additive to cosmetics to maintain moisture levels (Graf *et al*., 2008; Pastor *et al*., 2010; Reshetnikov *et al*., 2011) (Figure 4.5).

Hopene are pentacyclic triterpenoids produced that have important functions in many organisms, in *S. coelicolor* it is synthesised during formation of aerial hyphae to ease barrier stress on the biological membranes (Poralla *et al*., 2000). Pamamycins belong to a group of macrodiolide antibiotics produced by several *Streptomyces* species. They were reported to stimulate the formation of aerial mycelia in *Streptomyces alboniger* DSMZ40043 and inhibit the growth of Gram-positive bacteria and fungi (Rebets *et al*., 2015).

4.4 Gene clusters with high similarity to known gene clusters

As well as clusters that had 100% identity to known natural products, other clusters showed less similarity to known clusters. The most interesting of these natural products are mirubactin for gaining a deeper understanding of NRPS-catalysed siderophore biosynthesis and polyoxypeptin for its potent apoptosis inducing activity.

4.4.1 Mirubactin-like gene cluster

Mirubactin, a siderophore, was the first natural product discovered from *Actinosynnema mirum* and the first siderophore isolated from the genus *Actinosynnema*. Isolation, structure elucidation, and biosynthesis of mirubactin was reported by Giessen *et al*. in 2012 (Figure 4.6A). It was proposed that the gene cluster responsible for mirubactin biosynthesis covers approximately 29 kbp and contains 15 protein coding sequences (Figure 4.6B). Muribactin contains an unusual O-acyl

hydroxamate group, making muribactin biosynthesis a highly interesting system for detailed biochemical investigations (Figure 4.6C) (Giessen *et al*., 2012).

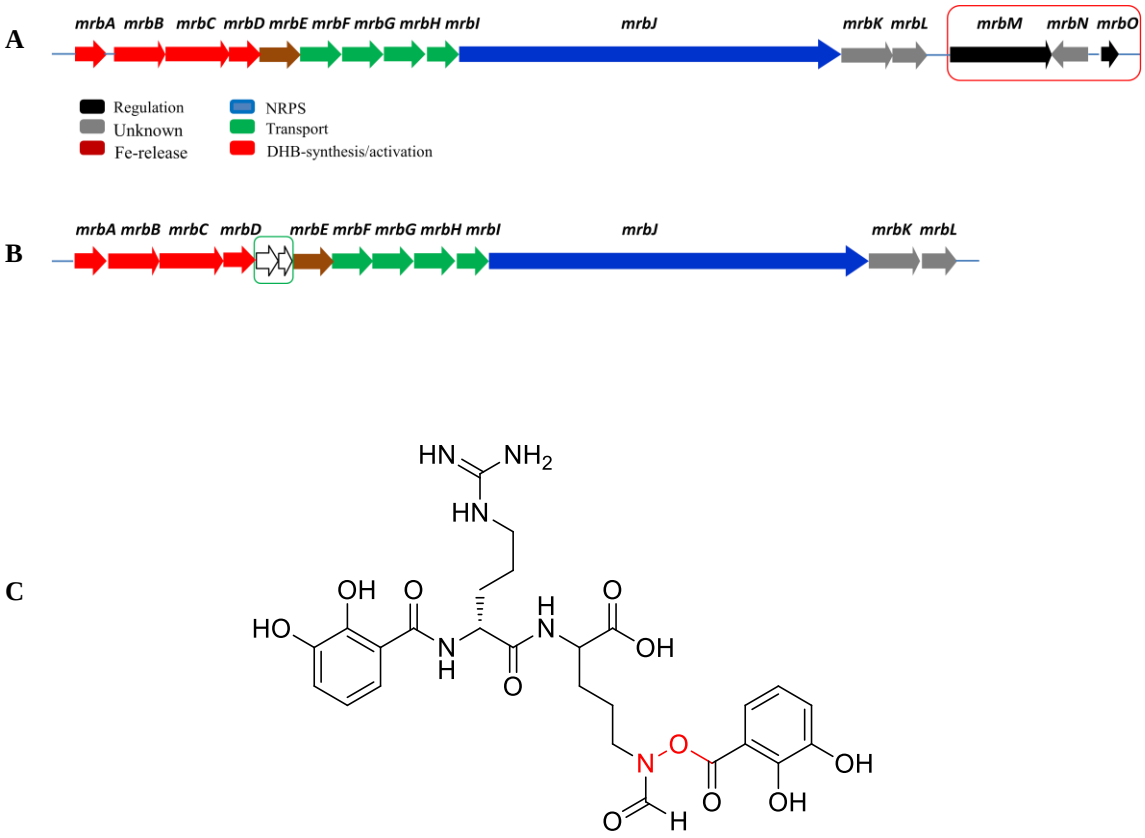


Figure 4.6 A. Muribactin biosynthetic gene cluster from *Actinosynnema mirum*. B. Muribactin biosynthetic gene cluster from *S. longispororuber* (cluster 8, 2161361 – 2219220 nt). C. Muribactin structure showing the acyl hydroxamic acid ester group.

antiSMASH analysis revealed that cluster 8 (2161361 – 2219220 nt) has 78% similarity with the gene cluster that is proposed to direct the biosynthesis of mirubactin (Figure 4.6A). The cluster appears not to contain the *mrbM*, *mrbN* and *mrbO* genes. *mrbM* and *mrbO* are proposed to be responsible for the regulation of mirubactin biosynthesis. On the other hand, there are two genes between *mrbC* and *mrbD* that appear to encode a 2, 3-dihydro-2, 3-dihydroxybenzoate synthetase and an acyl carrier protein. This means that *S. longispororuber* might be capable of producing mirubactin-like siderophores under iron-limited growth conditions.

4.4.2 Polyoxypeptin gene cluster

Polyoxypeptin A was isolated from *Streptomyces* sp. MK498-98 F14. The polyoxypeptin A core consists of six unusual amino acid residues, including 3-hydroxyleucine, piperazic acid, N-hydroxyalanine, 5-hydroxypiperazic acid, 3-hydroxy - 3-methylproline, and N-hydroxyvaline at high oxidation states (Du *et al*., 2014).

The polyoxypetin biosynthetic gene cluster contains 37 protein coding sequences (CDSs) (Figure 4.7A) (Du *et al*., 2014). Structural analysis of polyoxypeptin led to the hypothesis that polyoxypeptins are assembled by a hybrid PKS/NRPS system. Within the polyoxypeptin biosynthetic gene cluster, 4 type I modular PKS genes (*plyTUVW*), encoding four PKS modules, and 4 modular NRPS genes (*plyXFGH*), encoding 6 NRPS modules, are responsible for the assembly of the polyoxypeptin core structure (Figure 4.7C).

The compounds in the azinothricin family, which polyoxypeptin A belongs to, show diverse biological activities, such as potent antibacterial, antitumor (Agatsuma *et al*., 1997) and anti-inflammatory activities (Tsuji *et al*., 1995). Both polyoxypeptin A and polyoxypeptin B were confirmed to be potent apoptosis-inducing compounds (Umezawa *et al*., 1999).

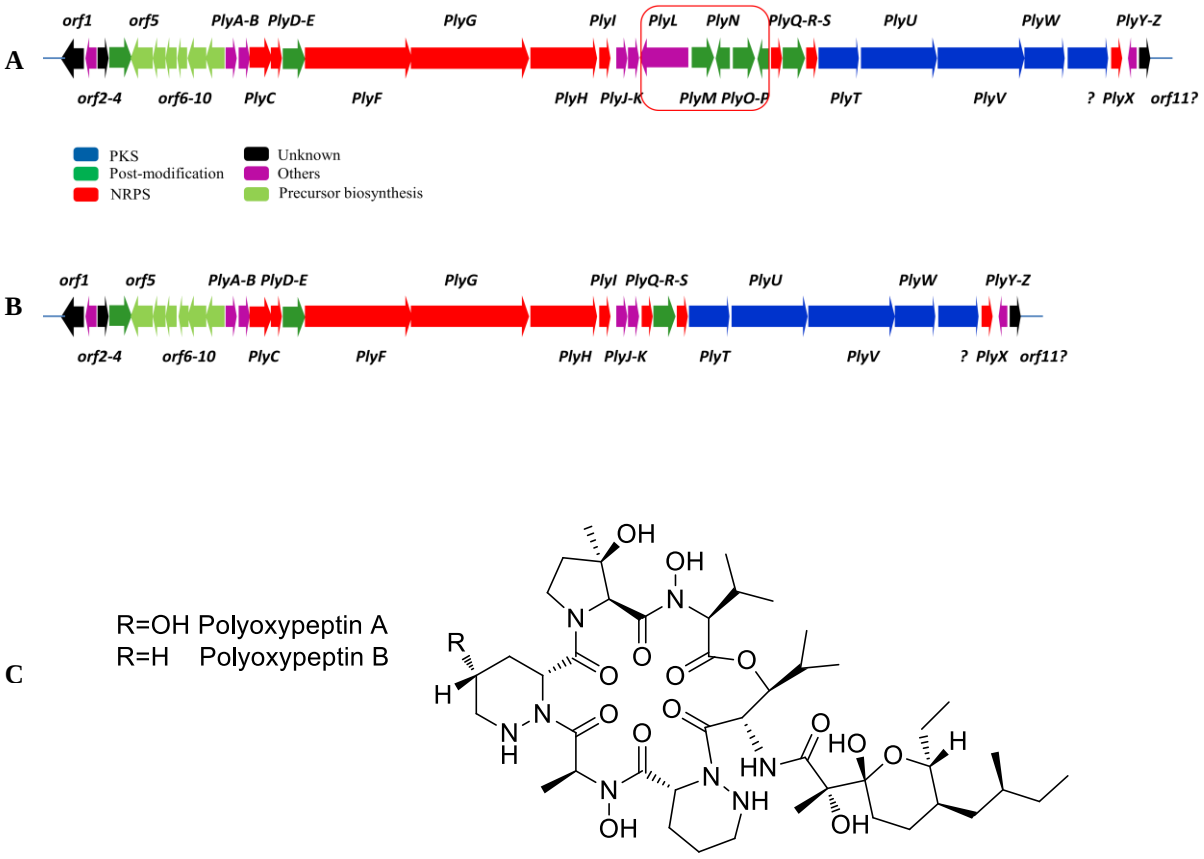


Figure 4.7 A. Organisation of polyoxypeptin biosynthetic gene cluster from *Streptomyces* sp. MK498-98 F14 B. Organisation of polyoxypeptin-like biosynthetic gene cluster from *S. longispororuber* (cluster 22, 4593178–4711093 nt). C. polyoxypeptin structure.

Cluster 22 (4593178 – 4711093nt) showed 67% similarity with a gene cluster that is proposed to direct the biosynthesis of polyoxypeptin (Table 4.1).

The cluster appears not to contain *plyL*, a putative regulatory gene and the *plyM*, *plyN*, *plyO* and *plyP* genes, proposed to be responsible for catalysing post assembly line modifications (Figure 4.7A and B). It is therefore possible that this cluster directs the biosynthesis of a polyoxypeptin analogue.

4.5 Cryptic gene clusters

Two clusters were identified, cluster 15 and cluster 24 (Table 4.1), which showed low similarity to previously characterised gene clusters, and are therefore predicted to encode novel natural products or natural products for which the gene cluster has not yet been identified. Since these gene clusters largely consist of PKS, NRPS or PKS/NRPS genes the product of the gene clusters can be predicted. The clusters were therefore analysed by sequence comparison and, in collaboration with Dr Matthew Jenner, Dr Douglas Roberts and Dr Lona Alkhalaf, the structures of the metabolic product of each of the assembly lines was predicted.

4.5.1 Cluster 15

Analysis of cluster 15 identified by antiSMASH, which shows 53% similarity to the gene cluster that directs the biosynthesis of ECO-02301, suggested it directs the production of a novel metabolite (Figure 4.8). ECO-02301 produced by *S. aizunensis* NRRL B-11277 demonstrated antifungal activity against a variety of fungi that are pathogenic to humans (McAlpine *et al*., 2004). This similarity is likely only due to the fact that they both encode large type I modular polyketide synthase assembly lines, as the predicted product of the PKS encoded by cluster 15 does not bear a significant resemblance to ECO-02301 (see below). Cluster 15 also contains genes encoding an NADPH: quinone reductase, a glycosyl hydrolase, and an amine oxidase, in addition to a few regulatory genes.

Cluster 15 is proposed to contain eight genes encoding type I modular PKS subunits. The NCBI online database and the PKS/NRPS online server (nrps.igs.umaryland.edu) were used to analyse the domains of the PKS. Sequence comparison of the acyltransferase (AT) domains with AT domains of known specifity was used to

predict the substrate specifity of each AT domain in the assembly line. Specificity for methylmalonyl-CoA is predicted by the presence of both a YASH motif and a GHSxG motif. Alternatively, the presence of a HAFH motif and a branched non-polar residue within the conserved GHSxG motif is indicative of incorporation of malonyl Co-A (Haydock *et al*., 1995; Del Vecchio *et al*., 2003).

The activity of the acyl carrier protein (ACP) within each PKS was determined by the presence of a conserved serine residue and the activity of ketosynthase (KS) domains was determined by the presence of a CHH catalytic traid. By this definition the KS domain of module 10, encoded by gene number 4, has been assigned as inactive (Figure 4.8).

The activity and stereospecifity of the ketoreductase (KR) domains was analysed by sequence comparisons with KR domains of known stereospecifity. The presence of a conserved LDD indicated the formation of an A-type or B-type (R)-configured β-hydroxyl group. The presence or absence of a proline residue in the catalytic region indicated (R) or (S)–stereochemistry, respectively (Caffrey 2003; Keatinge-Clay 2007). The functionality of the KR domain was assessed by looking for the presence of conserved tyrosine and asparagine residues.

The activity of the dehydratase (DH) domains was predicted by searching for four conserved motifs (HxxxGxxxxP, GYxYGPxF, LPFxW and DxxxQ/H). Any DH domains with motifs that deviated significantly were assigned as inactive (Keatinge-Clay 2008). For DH domains acting on β-hydroxyl groups assigned to have (*R*)-stereochemistry the double bonds in the product were predicted to have *trans*geometry. Conversely, for β-hydroxyl groups assigned to have (*S*)-

stereochemistry the corresponding double bonds in the product were predicted to have *cis*-geometry.

Sequence comparison of the enoyl reductase (ER) domains with ER domains of known sereospecifity predicted their stereospecifity. A conserved alanine (occasionally valine or phenylalanine) indicated (*S*)-stereochemistry while a conserved tyrosine denoted (R)-stereochemistry (Kwan *et al*., 2001; Kwan and Schulz 2001).

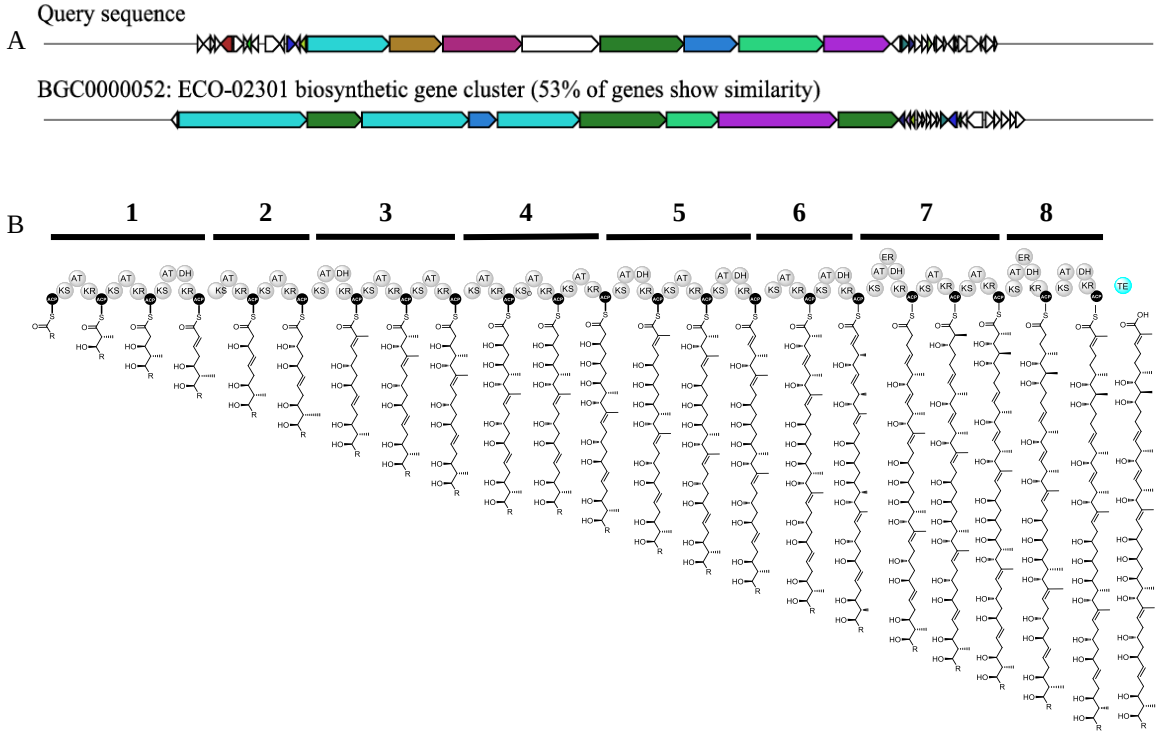


Figure 4.8 A. Comparison of the organisation of gene cluster 15 from *S. longispororuber* with the ECO-02301 biosynthetic gene cluster (the most similar cluster for which a metabolic product is known). B. Predicted subunit and domain organization of the type I modular PKS encoded by cluster 15 of *S. longispororuber* and the structure of the novel polyketide chain it is hypothesized to assemble.

4.5.2 Cluster 24

Cluster 24 (4754119 - 4864371nt), which encodes a hybrid *trans*-AT PKS/NRPS, showed low levels of similarity to clusters that direct the production of known metabolites (Figure 4.9 A). The cluster appears to contain five megasynth(et)ase-encoding genes; two of them encode trans-AT PKS subunits, two encode NRPS subunits and one encodes a hybrid trans-AT PKS/NRPS subunit. The substrate specifities of each domain were predicted according to literature methods, (Stachelhaus *et al.*, 1999; Challis *et al.*, 2000; Nguyen *et al.*, 2008) as described for cluster 15, above.

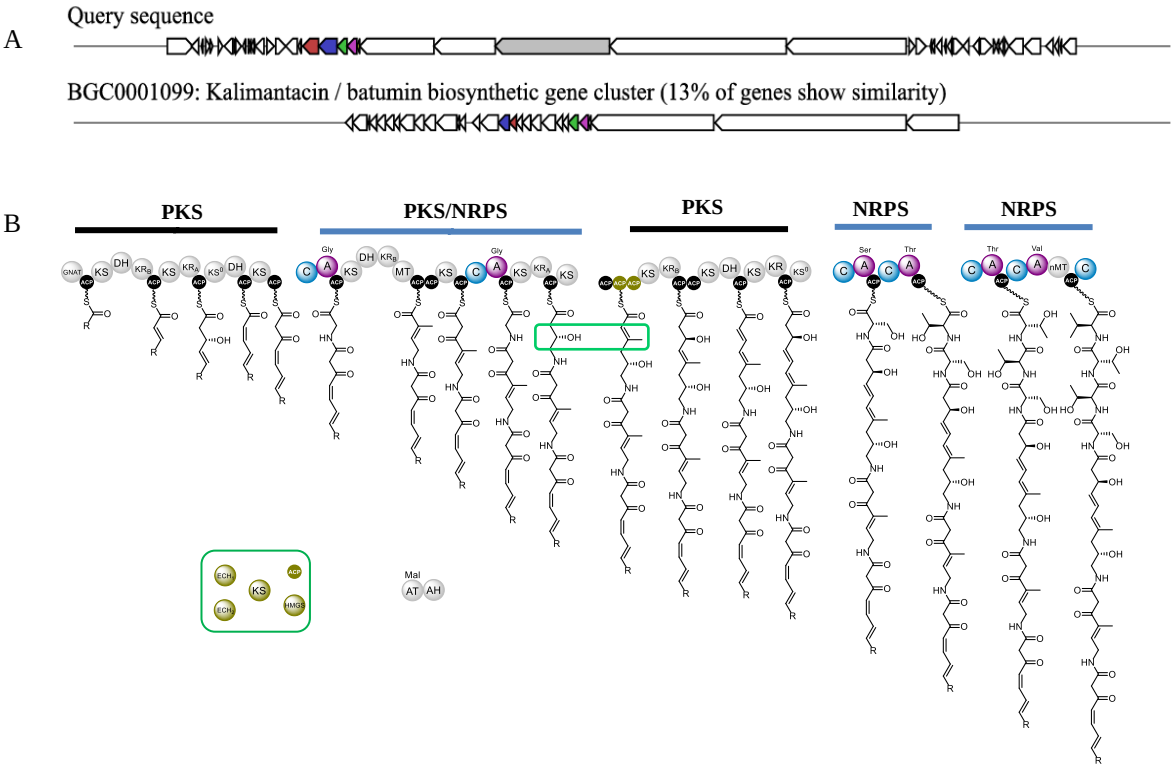


Figure 4.9 A. Comparison of the organization of gene cluster 24 in *S. longispororuber* with the kalamanticin biosynthetic gene cluster (the most similar cluster of known function). B. Proposed module and domain organization of the PKS/NRPS assembly line encoded by cluster 24, which is hypothesised to direct the biosynthesis a novel metabolite. The β-methyl branch-installing enzymes highlighted by the green box are proposed to act on the intermediate attached to the ACP domains in green.

In additon, the cluster contains genes encoding enoyl-CoA hydratase (ECH_1, ECH_2), ketosynthase (KS) and 3-hydroxyl-3-methylglutaryl-CoA synthase (HMGS) ezymes known to be involved in the installation of β-methyl branches during *trans*-AT PKS-mediated chain assembly. These enzymes are proposed to introduce a β-methyl branch into the intermediate attached to the ACP domains highlighted in green (Figure 4.9 B and Figure 4.10) (Buchholz *et al.*, 2010).

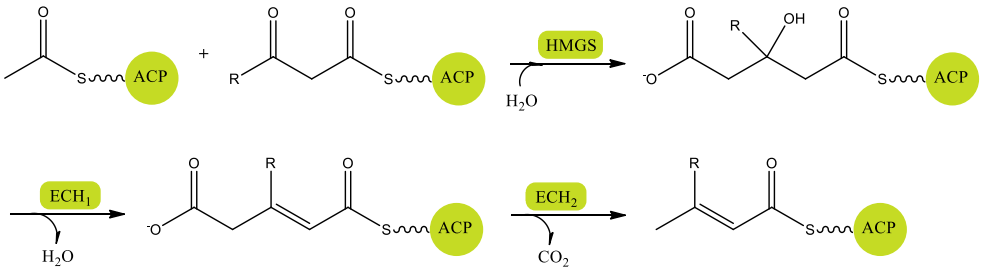


Figure 4.10 Installation of polyketide β-branching by HMGS.

4.6 Metabolites profile of *S. longispororuber*

To investigate the metabolite production profile of *S. longispororuber*, the strain was first grown on a selection of solid media including R5, SMMR and Isp2 (Figure 4.11). The mycelia were separately extracted with ethyl acetate and methanol.

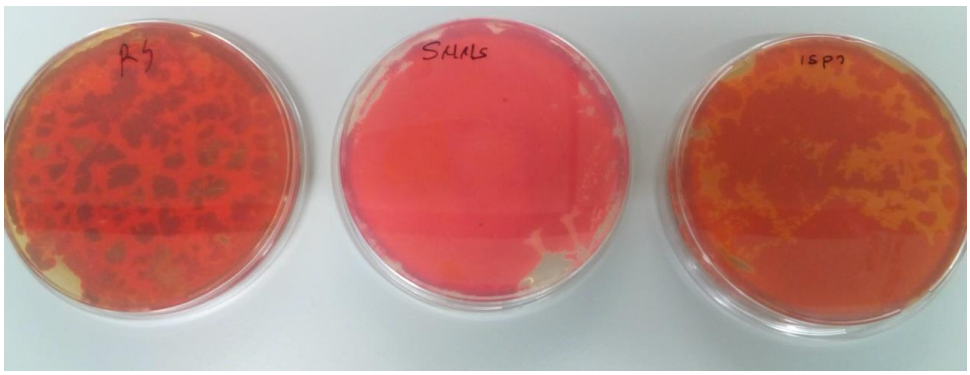


Figure 4.11 *S. longispororuber* grown on R5, SMMS and Isp2 solid media.

The extacts were then analysed by LC-MS to determine if any of the natural products proposed to be encoded within the *S. longispororuber* genome could be detected. Unfortunatey, due to the large amount of red-pigmented prodiginines produced, no other compounds could be detected (Figure 4.12). In order to try and detect the predicted structures, extracted ion chromatograms were used to search for the production of compounds with *m/z* values corresponding to the predicted masses of the products of the gene clusters identified via analysis of the genome sequence.

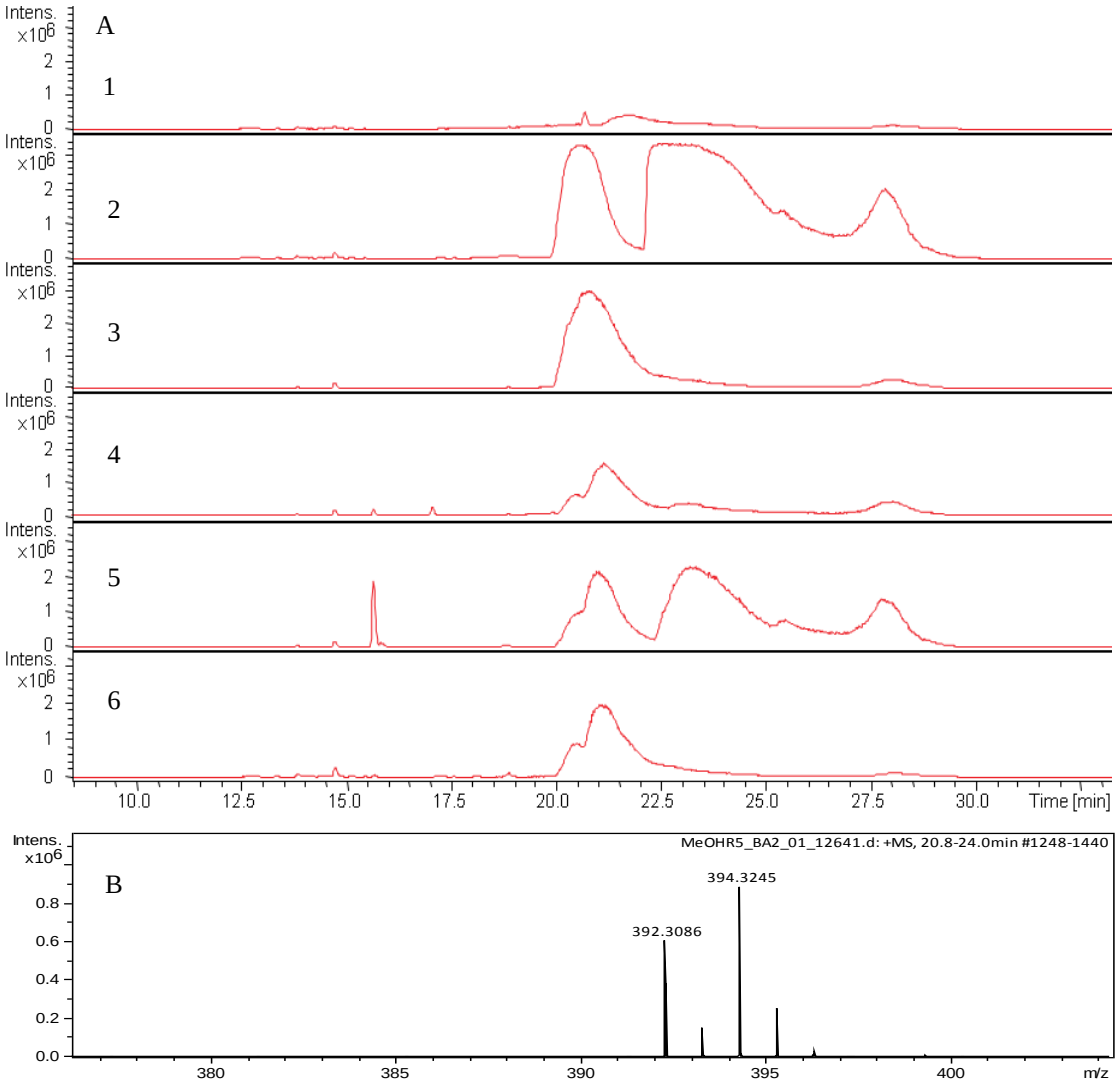


Figure 4.12 A. Base peak chromatograms from LC-MS analysis of organic extratcs of *S. longispororuber* grown on different solid media. 1. Isp2- EtOAc, 2. R5- EtOAc, 3. SMM- EtOAc, 4. Isp2-MeOH, 5. R5-MeOH and 6) SMM-MeOH. B. High resolution mass spectrum confirming that compounds with molecular formulae corresponding to undecylprodigiosin and metacycloprodigiosin are present in the extracts.

These extracted ion chromatograms showed no peaks corresponding to the molecular weights of muribactin, polyoxypeptin, or the predicted metabolic products of clusters 15 or 24 (Figure 4.12).

S. longispororuber was then grown in 50 ml of liquid media (R5, SMM and Isp2). The strain did not grow in Isp2 media so mycelia extraction was only carried out following growth in R5 and SMM, separately using ethyl acetate and methanol. Unfortunately, as with solid media, no products other than the prodiginines could be detected (Figure 4.13).

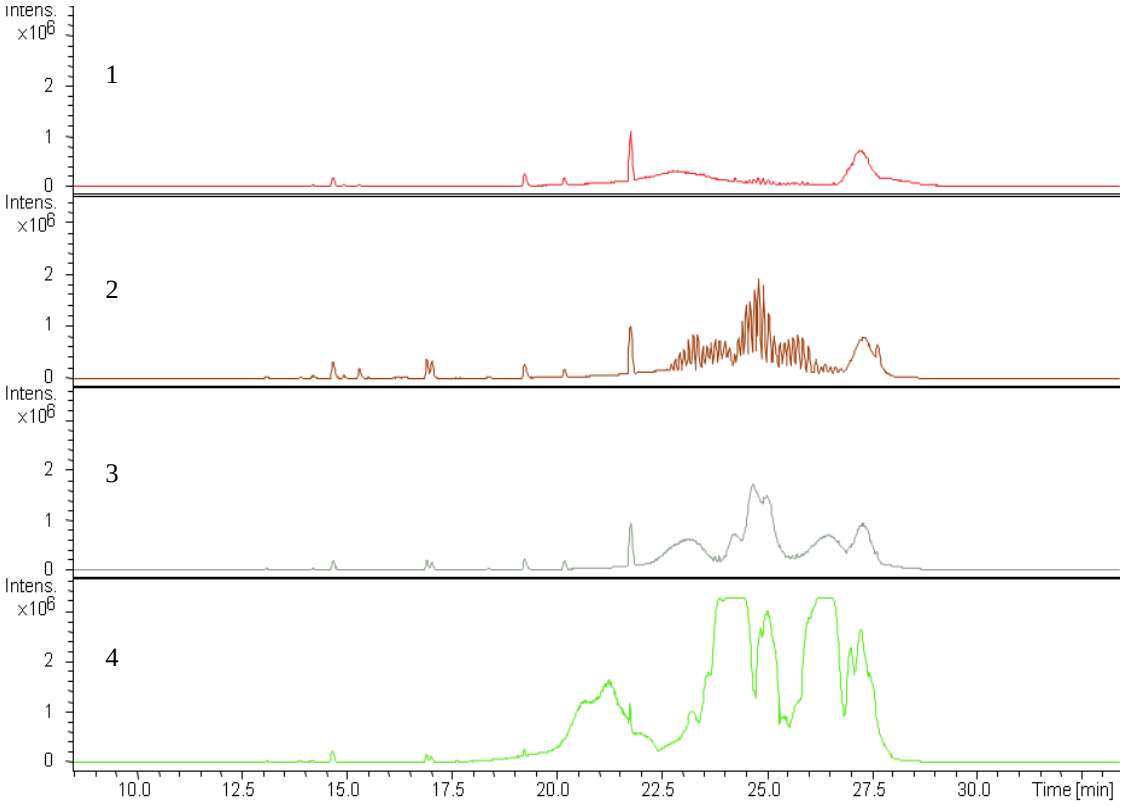


Figure 4.13 Base peak chromatograms from LC-MS analysis of extracts of *S. longispororuber* grown in R5 or SMM liquid media. 1. R5- EtOAc, 2. R5-MeOH. 3. SMM- EtOAc and 4. SMM-MeOH.

It is possible that the very large amount of undecylprodigiosin and metacycloprodigiosin produced is interfering with the detection of other metabolites, or that the high production levels of undecylprodigiosin and metacycloprodigiosin is

consuming the precursors required for the biosynthesis of the other metabolites. Gene cluster activation, deletion of the metacycloprodigiosn biosynthetic gene cluster or specific cultivation conditions might be required to produce these metabolites from *S. longispororuber*.

In conclusion, in addition to correcting the *mcpG* sequence, the genome sequence of *S. longispororuber* revealed that, apart from the metacycloprodigiosin biosynthetic gene cluster, the strain contains several putative natural product biosynthetic gene clusters. Among these clusters, seven showed 100% similarity to known gene clusters from other *Streptomyces* species. Others showed less similarity and in some cases are predicted to direct the biosynthesis of novel metabolites.

Conclusions and Future work

5. Conclusions and Future work

5.1 Conclusions

S. longispororuber redG orthologue, *mcpG*, was heterologously expressed in *S. albus* and *S. coelicolor* M511 and fed with undecylprodigiosin (**2**) extracted from *S. coelicolor* W31 (M511Δ*redG::scar*; which produces **2** but not streptorubin B (**3**)). Metacycloprodigiosin was observed in the *S. albus* host strain but the amount of conversion was low. It was hypothesised that a catalytically important complex could be formed between *mcpG* and *mcpH in vivo* as when *redH* and *redG* were expressed together more streptorubin B was produced (Sydor, 2010).

mcpH was therefore heterologously expressed in *S. albus* and fed with synthetic 2-UP (**15)** and MBC (**16)** and a large amount of undecylprodigiosin (**2**) was produced. This confirmed the role of McpH as an enzyme that catalyses the condensation of **15** and **16** to form **2** in *S. longispororuber*. Heterologous expression of *S. longispororuber mcpHG* in *S. albus* and feeding with synthetic **15** and **16** resulted in the production of undecylprodigiosin and a small amount of metacycloprodigiosin, but the quantity of metacycloprodigiosin was not significantly higher than in the experiments employing *mcpG* alone.

The construct used by Paulina Sydor in the original McpG proof of function experiments was shown to encode a truncated protein due to a sequencing error. The correct sequence of *mcpG* was obtained via whole genome sequencing and expressed both separately and together with *mcpH*. Repeating the experiments described above with these new constructs did not result in improved production levels for metacycloprodigiosin.

Following the discovery that the construct previously used by Paulina Sydor to determine McpG activity is truncated, it was decided to repeat her experiments using the full length gene. Heterologous expression of full length *mcpG* in *S. coelicolor* W31(a Δ*redG::scar* mutant of the M511 strain), which produces undecylprodigiosin but not streptorubin B resulted in production of a carbocyclic undecylprodigiosin derivative. This metabolite was identified as metacycloprodigiosin by HRMS and NMR spectroscopic analysis.

Surprisingly, expressing *mcpG* in *S. coelicolor* W31 resulted in higher levels of metacycloprodigiosin production than expression feeding of undecylprodigiosin to *S. albus* expressing *mcpHG*. It is possible that the *mcpG* gene is toxic to *S. albus* in the absence of its substrate and thus the bacteria rapidly select for variants that are non-functional. These results confirmed that McpG catalyses an oxidative cyclisation reaction to form a C-C bond between C-4 and C-9 of undecylprodigiosin to yield metacycloprodigiosin in *S. longispororuber*.

The *rph* gene cluster in *S. griseoviridis* which consists of 21 genes that are homologous to genes in the red cluster of *S. coelicolor* contains four *redG* orthologues (*rphG*, *rphG2*, *rphG3* and *rphG4*). Three of these (*rphG*, *rphG2* and *rphG4*) contain the conserved [2Fe-2S] cluster-binding motif characteristic of Rieske oxygenases. The cluster has been proposed to be responsible for the biosynthesis of prodigiosin R1 (**20**) and roseophilin (**21**). Comparison of the structures of roseophilin and prodigiosin R1 with their putative common precursor 11-methyldodecylprodigiosin suggests that RphG, RphG2 and RphG4 are each required to form one C-C bond (Hayakawa *et al*., 2009).

Expression of *rphG*, *rphG2*, *rphG3* and *rphG4* in *S. albus* and feeding with synthetic 11-methyldodecylprodigiosin did not result in the production of any carbocyclised derivatives. The four genes were thus expressed in *S. coelicolor* W31 this also did not yield any carbocyclic derivatives of **2**, possibly because undecylprodigiosin is not a substrate for the RphG enzymes. The sequences used to create the *rphG* expression constructs were those originally reported by Hayakawa *et al.* Towards the end of the project it was reported that the original sequence for *rphG* was incorrect. Expression of the correctd *rphG* sequence (Hayakawa *et al*., 2009; Kimata *et al*., 2016) in *S. coelicolor* W31 did not result in the production of any carbocyclic undecylprodigiosin derivatives, even when a universal RBS (AGGAGG) was used.

To investigate whether a complex is made between RphH and the RphG, RphG2, RphG3 and RphG4 enzymes, and to overcome the potential problem of high substrate specificity, *rphH* was synthesised to be expressed in *S. albus* both alone and in combination with each of *rphG*, *rphG2*, *rphG3* and *rphG4*. Expression of *rphH* alone in *S. albus* and feeding with synthetic 11-methyldodecylpyrrole (**25**) and MBC (**16**) did not give any condensation product suggesting it is not active, or not being expressed under the conditions used. As sequence alignment of RphH with RedH and McpH showed 139 extra amino acids in the N-terminus of the protein, two truncated variants of RphH were investigated, but neither showed any catalytic activity.

The complete genome sequence of *S. longispororuber* (9,239,768 bp) revealed many cryptic natural product biosynthetic gene clusters. Using antiSMASH, around 30 gene clusters were predicted to be responsible for the production of different natural products. Six clusters showed a high degree of similarity to known gene clusters in other Actinobacteria. These are predicted to direct the production of albaflavenone, coelichelin, 2-methylisoborneol, and ectoine. Two others were similar, but not

identical to known gene clusters and were predicted to direct the production of mirubactin and a polyoxypeptin derivative. Two of the gene clusters were predicted to direct the production of novel metabolites.

5.2 Future work

The stereochemical course of the McpG-catalysed reaction could be elucidated by feeding 2-undecylpyrrole stereoselectively deuterium labelled at C-9' to a mutant of *S. longispororuber* in which the *redLK* homologue has been deleted. However, the genetic tractability of *S. longispororuber* has previously been investigated by Paulina Sydor and was found to be poor. An alternative approach could involve cloning of the entire *mcp* cluster directly from the genome of *S. longisporuber* using transformation-associated recombination in yeast. Chuan Huang, a research fellow in the Challis group, has developed an efficient method for such direct gene cluster cloning, based on the pCAP01 and pCAP03 vectors reported by Moore and co-workers (Yamanaka *et al*., 2014). If expression in a heterologous host, e.g. *Streptomyces albus*, leads to production of metacycloprodigiosin the *redLK* homologue could be deleted in the expression construct and the stereoselectively deuterium labelled 2-undecylpyrroles could be feed to the resulting mutant. The results of these experiments will be interesting to compare with those from the incorporation of 2-undecylpyrrole stereoselectively deuterium labelled at C-7' into streptorubin B via feeding to a *redL* mutant of *S. coelicolor*, because metacycloprodigiosin and streptorubin B have opposite absolute configurations at C-9' and C-7', respectively.

The roles played by RphG, RphG2 RphG3 and RphG4 in the biosynthesis of roseophilin and prodigiosin R1 merit further investigation. One approach to this could involve the cloning and heterologous expression of the *rph* gene cluster, using similar

methods to those proposed above for the *mcp* cluster. If this proved successful, *rphG*, *rphG2* *rphG3* and *rphG4* could be independently deleted and the metabolites accumulating in each strain could be characterised. However, this would require access to genomic DNA from *S. griseoviridis*.

Analysis of the *S. longispororuber* genome sequence identified two gene clusters that seem likely to direct the production of novel metabolites. The approach discussed above could also be employed to clone and heterologously express these gene clusters and rational methods for inducing their expression could be explored (Rutledge and Challis, 2015). This may ultimately allow the metabolic products of these gene clusters to structurally characterised and investigated for interesting biological activity.

Materials and Methods

6. Materials and Methods

6.1 Enzymes, media, chemicals and equipment

Restriction enzymes and enzymes used to modify DNA were provided by Thermo Scientific and New England Biolabs (USA). Kits to extract DNA from gels and to isolate plasmid DNA from *E. coli* were obtained from Thermo Scientific, Qiagen (USA). HPLC grade methanol and water was purchased from either Fisher Scientific or Sigma Aldrich. HPLC grade formic acid was purchased from Sigma Aldrich and used within one month of opening.

LB broth medium was provided by Fisher Bio Reagents. Most of the ingredients for growth were obtained from Difco and Becton, Dickinson and co. The media used (LB, R5, SFM, SMM and Isp2) were made according to manufacturers guidelines or from known procedures. Chemicals were bought from Difco (USA), Sigma-Aldrich (USA), Thermo Scientific, Fisher Scientific (UK).

PCRs were performed in an Eppendorf Master cycler Personal. A BioRad Gene Pulsar II was used for electroporation in conjunction with the BioRad Pulse Controller Plus. To analyse prodiginine production LC-MS was used (Bruker MaXis Impact Mass Spectrometer). The optical density of cells was checked by Beckman coulter DU7400 Spectrophotometer. To measure DNA concentrations a nano-drop ND-1000 spectrophotometer was used. Optical density of microbial cultures was measured using a Beckman coulter DU7400 Spectrophotometer. Bassaire laminar flow hood was used to handling *Streptomyces*. To purify secondary metabolites an Agilent 1200 HPLC instrument was used. NMR data was acquired from Bruker Advance spectrometer, 400 MHz instrument.

6.1.1 Strains and plasmids

The strains and plasmids used in this study are listed in Table 5.1.

Table 6.1 Strains and plasmids.

Bacterial Strains	**Genotype**
E.coli ET12567/pUZ8002	*dam*, *dcm*, *hsdM*, *hsdS*, *hsdR*, *cat*, *tet*; plasmid pUZ8002*: tra,* KanR, *RP4 23;* Genome*: dam*, *dcm*, *hsdS*, CmR, TetR
E. coli TOP10	F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G
***Streptomyces* Strains**	**Genotype**
M511	*S. coelicolor* A(3)2, SCP1$^-$, SCP2$^-$ act$^-$ (ΔactII-orf4)
M511 W31	*S. coelicolor* M511Δ*redG*::*scar*
S. longispororuber	Wild type, lab Stock
S. albus J1704	Wild type, lab Stock
***Transconjugant Streptomyces* Strains**	
Plasmids and fosmids	
pOSV556	OriT, *ErmE* promotor, int-pSAM2, AmpR, hygR, CoIE1 Ori
Fosmid 3G3	Chl R. lab stock by Paulina Sydor.
pOSV556*mcpG*-tr	*mcpG* from *Streptomyces longispororuber*, tr = truncate
pOSV556*mcpHG*-tr	*mcpHG* from *Streptomyces longispororuber*, tr = truncate
pOSV556*mcpG*	*mcpG* from *Streptomyces longispororuber*
pOSV556*mcpHG*	*mcpHG* from *Streptomyces longispororuber*
pOSV556*rphH*	*rphH* from *Streptomyces griseoviridis*
pOSV556 *rphH*-m1	*rphH* from *Streptomyces griseoviridis*, m1= mutant 1
pOSV556 *rphH*-m2	*rphH* from *Streptomyces griseoviridis*, m2= mutant 2
pOSV556*rphG*	*rphG* from *Streptomyces griseoviridis*
pOSV556*rphG2*	*rphG2* from *Streptomyces griseoviridis*

pOSV556*rphG3*	*rphG3* from *Streptomyces griseoviridis*
pOSV556*rphG4*	*rphG4* from *Streptomyces griseoviridis*

6.1.2 Primers

The primers used in this study for PCR amplification and sequencing are listed in Table 6.2.

Table 6.2 PCR amplification and sequencing primers

Name	Sequence 5'→ 3'	Characteristic
Seq1	TCCCGGCCCGCGCCACCA	Sequencing of *mcpH* and *mcpHG*
Seq2	AGCACGACACGGTCCTCG	
Seq3	CCACCCGGCCCCGCCCCG	
Seq4	CCGCATGGGCCGCCGCG	
Seq5	AGAAACTGCCGCTGCCG	
Seq6	AGTGGTATCCGATTCTGC	
RS1	GGAACAGGTCGAACAGGC	
RS2	GGAGAGCGAGGCGCATTT	
RS3	CCGAACTCACCCGCTACC	
RS4	CGCCTCTGTTCGCCATC	
RS5	CATTTCCGGAGCCGTTCG	Sequencing of *rphG*s
rphG2	CGCGGCGAACCGTACAAG	
mcpH-F	AAAGGGAAGCTTAGGAGGCCGCGTCATGACGA	PCR amplification of *mcpH*
mcpH-Rn	AAAGGGAGGCCTCTGCCTTCCCTGCTTCCTTG	
mcpH-F	AAAGGGAAGCTTAGGAGGCCGCGTCATGACGA	PCR amplification of *mcpHG*
mcpHG-R3n	TTTAGGCTTGGAGCCGTTCGTGCCGGTCG	
mcpHG-Fw	AAAGGGaagcttAGGAGGCCGCGTCATGAC	PCR amplification of corrected *mcpHG*
mcpHG-Rv	AAAGGGaggcctTCAGGGCGCGGTCTTGCT	
rphG3-F	AAAGGGAGGCCTAGGAGGGTCCGCGTGCGATGTTCCCCCA	PCR amplificatio of *rphG3*
rphG3-R	AGGGCGAATTGGGTACCG	

RS1 new	CGCAGGAACAGGTCGAACAG	Sequencing of *mcpH* and *mcpHG*
RS4 new	GTCGGCGAAGTCCAGGAAGG	
RphH-F	TTTCACGTGAGGAGGCCCGCCCGTGTCTGTGC	PCR amplification of *rphH*
RphH-R	CCCAAGCTTTCACGCACCACCCCCGGCTGCG	
rphH-F-mu1	TTTCACGTGAGGAGGGTCCGCATGGACACGCTGGGGGGCAAGG	
rphH-F-mu2	TTTCACGTGAGGAGGCGTCGGCGGACGTC	
rphH-R-mu	CCCAAGCTTTCACGCACCAC	
rphG1-F	TTTAAGCTTGTACACCGTGCCGCACGCACAG	PCR amplification of *rphG1*
rphG1-R	TTTCTCGAGGTCATGCGGAGGCCGCGATGTC	
rphG1-F-RBS	TTTAAGCTTAGGAGGCAGAGAAATGATCCCGAATCAGTG	
rphG1-R2	TTTCTCGAGTCATGCGGAGGCCGCGATGTC	
rphG-S2fw	TGGACGAGCTGAACGCGAAG	Sequencing of *rphG1*
rphG-S1rv	ATCCAGATGAGCCCGAGGTC	

6.1.3 Antibiotics

Antibiotic stock solutions were prepared by dissolving in the appropriate solvent and sterilised by filtration through a 0.2 um filter (Table 6.3) (Sambrook and Russell, 2001).

Table 6.3 Antibiotics stock solutions.

Antibiotic	**Stock sol. (mg/ml)**	***E. coli* (µg/ml)**	***S. coelicolor* (µg/ml) overlay**	***S. coelicolor* (mg/plate) overlay**
Ampicillin	50	100	-	-
Chloramphenicol	25	25	-	-
Hygromycin	50	50	25	1.25
Kanamycin	50	50	100	5.0
Nalidixic acid	25	25	20	1.0

6.1.4 Culture media

The media for culturing *Streptomyces* and *E. coli* were prepare according to published procedures as described by Kieser *et al*. (2000) and all media were autoclaved directly after preparation.

6.1.4.1 Liquid media

Recipes for liquid media used are as follows (Table 6.4).

Table 6.4 Liquid media.

LB Broth		**TSB (Tryptone Soya Broth)**	
Luria-Bertani powder	25 g/L	Oxoid Tryptone Soya Broth powder	30g
H_2O	to 1L	H_2O	to 1L

6.1.4.2 Solid media

Recipes for solid media used are as follows (Table 6.5).

Table 6.5 Solid media.

LB broth		**SFM (Mannitol soya flour medium)**	
Luria-Bertani powder	25 g/L	Soya flour	20g
Difco Bacto Agar	15 g	Mannitol	20g
H_2O	to 1L	Difco Bacto Agar	20g
		H_2O	to 1L
R5 medium			
Ingredients (before autoclaving)		**Ingredients/1L (after autoclaving)**	
Sucrose	103 g	KH_2PO_4 (0.5%)	10 mL
K_2SO_4	0.25 g	$CaCl_2.2H_2O$ (5M	4 mL
$MgCl_2.6H_2O$	10.12 g	L-proline (20%)	15 mL
Glucose	10 g	NaOH (1N)	7 mL
Difco Casaminoacids	1g	Trace Element Solution	2 mL
Difco Yeast Extract	5 g		

TES Buffer	5.73 g		
Agar	22 g		
H_2O	to 1L		
Liquid R5 medium was used using the same ingredients except Agar. All "Before autoclaving ingredients" were mixed and medium was autoclaved. At the time of use the medium was melted and then "After autoclaving ingridents" were added			
Supplement Minimal Medium (SMM)		**Trace Element Solution**	
$MgSO_4.7H_2O$ (0.2M)	25 mL	$ZnSO_4.7H_2O$	0.1 g/L
TES Buffer (0.25M, pH 7.2)	100 mL	$FeSO_4.7H_2O$	0.1 g/L
$NaH_2PO_4 + K_2HPO_4$ (50 mM each)	10 mL	$MnCl_2.4H_2O$	0.1 g/L
Trace Element Solution	1 mL	$CaCl_2.6H_2O$	0.1 g/L
Casaminoacids (20%)	10 mL	NaCl	0.1 g/L
H_2O	804 mL	H_2O	to 1 L
Glucose (20%)	50 mL		
All solutions were autoclaved or filtered separately and added to sterile H_2O		Filtered	
Isp2			
Difco Yeast Extract	4g	Liquid Isp2 medium was prepared using the same ingredients without addition of agar	
Difco Malt Extract	10g		
Difco Dextrose	4g		
Agar	20g		
H_2O	to 1L		

6.1.5 PCR

Reaction conditions were as follows:

1) Denaturation: 94°C, 2 min

2) Denaturation: 94°C, 45 sec

3) Primer annealing: (variable) °C, 45 sec

1, 2 and 3: 35 cycles

4) Extension: 72°C, 2 sec.

For Phusion polymerase enzyme the denaturation temperature was 98°C.

PCRs were performed using the following reaction mixture (Table 6.6)

Table 6.6 The mixture of PCR reaction.

PCR reaction elements	amount
Primers (10µM)	1 µl each
DNA template	1 µl
Buffer (10x)	5 µl
dNTPs (5 mM)	2 µl each
DMSO (100%)	2.5 µl
DNA polymerase (2.5 U/µl)	1 µl
Distilled water	36.5 µl
Total volume	50 µl

6.2 Methods

6.2.1 Cloning of *mcpHG*, *mcpG* and *rphGs* genes

The genes were amplified by PCR according to standard procedures (section 6.1.5). After amplification each sample was run on a 1% agarose gel and purified from the gel using GenJET Gel Extraction Kit (Fermentas). Directly after purification, ligation of the insert with pOSV556 was carried out (Section 5.2.4). Half of the ligation reaction (5 µl) was then introduced to *E. coli* TOP10 cells by chemical transformation and, after incubation, cells with each construct were spread on two LB plates containing ampicillin and incubated overnight. For each construct plasmid DNA from about 20 single colonies were isolated and analysed by restriction digestion by suitable enzymes.

6.2.2 Agarose gel electrophoresis

Agarose gels were prepared and run in 1x TBE buffer. Depending on the size of the DNA fragments, 0.8%, 1.0% or 1.5% agarose gels were used. Electrophoresis was carried out for 1h at 100v. A 1kb DNA molecular size marker (Thermo Scientific) was used to estimate the size of the DNA molecules being analysed. Gel red (Biotium) was used to visualise DNA fragments.

6.2.3 Digestion of DNA with restriction enzymes

Restriction digestions were carried out according to the manufacturer's instructions (Thermo scientific). Restriction enzymes always occupied 10% of the total volume and 10% of an appropriate buffer (10x concentrated) was used.

6.2.4 Ligation of DNA

DNA to be ligated was purified using a DNA purification kit (Thermo) eluted in H_2O. In general, a molar ratio of vector: insert of 3:1 was used. Ligation reactions in 20 µl final volumes were carried out for 1h using Quick Ligation Kit (Thermo).

6.2.5 Growth, storage and manipulation of *E. coli*

The procedures described by Sambrook and Russell (2001) were used for culturing *E. coli*.

6.2.5.1 Growth conditions

E. coli was grown overnight on LB agar or in LB broth (180 rpm, 37 °C). Plasmid-containing cells were selected for using the appropriate antibiotic at the concentration given in Table 5.3.

6.2.5.2 Storage of Strains

For long term storage, overnight LB cultures of *Streptomyces* or *E. coli* were mixed with 70% glycerol to a final concentration of 20% and stored at -78 °C.

6.2.6 Preparation of electrocompetent *E. coli* cells

E. coli was used to inoculate 10 ml of LB and the resulting culture was grown overnight with shaking at 37 °C. 100µl of this preculture was inoculated into 10 ml LB and grown at 37 °C for 3-4 h shaking at 200 rpm to an OD_{600} of ~ 0.6. The cells were recovered by centrifugation at 3000 rpm for 7 min at 4 °C. After decanting the medium the pellet was resuspended by gentle mixing in 10 ml of ice-cold 10 % autoclaved glycerol. The cells were centrifuged as before and resuspended in 5ml of ice-cold 10 % autoclaved glycerol,

centrifuged, the supernatant was decanted and the cells were resuspend in the remaining ~ 100 µl of 10 % glycerol.

6.2.7 Preparation of chemically competent *E. coli* cells

E. coli was used to inoculate 10 ml of LB and the resulting culture was grown overnight with shaking at 37 °C. 100 µl of this pre-culture was inoculated into 10ml LB and grown at 37 °C for 3-4h shaking at 200 rpm to an OD_{600} of ~ 0.6. The cells were recovered by centrifugation at 3000 rpm for 7 min at 4 °C. After decanting the medium the pellet was resuspended by gentle mixing in 10 ml of ice-cold washing buffer (80 mM $MgCl_2$ and 20 mM $CaCl_2$). The cells were centrifuged as before. The medium was decanted and the tubes were inverted for 1 min to allow any traces of medium to drain away. The pellet was resuspended in 100 µl of ice-cold 0.1 M $CaCl_2$. 50 µl of cells were used for each transformation.

6.2.8 Transformation of electrocompetent *E. coli* cells

An appropriate amount of the electrocompetent cell suspension (100 µl) was mixed with ~ 100ng DNA per transformation. Electroporation was carried out in a 0.2 cm ice-cold electroporation cuvette using a BioRad GenePulser II set to: 200 Ω, 25 µF and 2.5 or 1.8 kV. The expected time constant is 4.5 – 4.9 ms. After electroporation, 1 ml ice-cold SOC (Super Optimal broth with Catabolite repression) was immediately added to the shocked cells and they were incubated with shaking for 1 h at 37 °C. Transformants were selected by spreading onto LB agar containing the appropriate antibiotic.

6.2.9 Transformation of chemically competent of *E. coli* cells

An appropriate amount of the suspension of the competent cells (50 µl) was transferred to an ice-cold 1.5 ml micro-centrifuged tube on ice. DNA was added (no more than 100 ng in a 10µl volume) and the suspension was gently mixed. The tube was stored on ice for 30min, transferred to a 42 °C water bath for 90 s and then put for 1min into ice. 200 µl of LB was added to the cells and they were incubated with shaking at 37 °C for 1h. Transformants were selected by plating on agar containing the appropriate antibiotics.

6.2.10 Growth, storage and manipulation of *Streptomyces*

The procedures used were adapted from those described by Kiesser *et al*. (2000).

6.2.10.1 Surface grown cultures for spore stock generation

To prepare spore stocks of *S. coelicolor, S. albus* and *S. longispororuber*, spore suspensions were prepared from confluent lawns growing on SFM agar medium. 3 mL of sterile water was added to the plates and the spores were suspended in the water by carefully scraping the surface of the mycelia with a sterile loop. The suspension was filtered through sterile non-absorbent cotton wool. The filtered suspension was centrifuged (10 min, 3000 rpm) to pellet the spores and the supernatant was poured off. The pellet was resuspended in the drop of water remaining in the tube and approximately the same volume of sterile 50% glycerol was added. Spores were stored at -20 °C.

6.2.10.2 Liquid grown cultures for genomic DNA isolation

Streptomyces cultures for genomic DNA isolation were grown in TSB medium for 48h at 30 °C and 180 rpm.

6.2.11 Conjugation of *E. coli* ET12567/pUZ8002 and *Streptomyces*

Conjugation was carried out following transformation of electrocompetent cells of *E. coli* ET12567/pUZ8002 with pOSV556/*rphG1*, *rphG2*, and *rphG4*, pOSV556/*mcpG* and pOSV556/*mcpHG*. Few colonies from each transformation were grown overnight at 37 °C in LB containing kanamycin (25µg/ml) and chloramphenicol (25 µg/ml) to maintain selection for pUZ8002 and the *dam* mutation, respectively, and ampicillin to maintain the plasmid. Then 100 µl of overnight culture was inoculated into 10 ml of fresh LB containing antibiotics as above and grown for ~ 4 h at 37 °C to an OD_{600} of 0.4. The cells were washed twice with 10 ml of LB to remove the antibiotic that might inhibit *Streptomyces* growth, and suspended in 1 ml of LB. In the meantime, for each conjugation 10 µl *Streptomyces* spores were added to 500 µl 2 × YT broth and heat shocked at 42 °C for 10 min. *S. albus* spores were additionally incubated at 30°C for 2.5 h. 0.5 ml of *E. coli* cell suspension was mixed with 0.5 ml of heat-shocked spores and the mixture was briefly centrifuged. Most of the supernatant was poured off, and the pellet was resuspended in the residual liquid and plated out on SFM agar containing 10 mM MgCl (without antibiotics) . The plates were incubated at 30 °C for 16-20 h.

Plates were overlaid with 1 ml of water containing 0.5 mg nalidixic acid (20 ul of 25 mg/ml stock; selectively kills *E. coli*) and 1.25 mg hygromycin (25 ul of 50 mg/ml stock) (Table 5.3) and incubated at 30 °C for further 3 days. Single colonies were then transferred to fresh SFM plates containing nalidixic acid and hygromycin.

6.2.12 Isolation and manipulation of DNA

6.2.12.1 Genomic DNA isolation from *S. coelicolor* and *S. longispororuber*

Streptomyces was grown in a 50 mL culture in TSB liquid medium for 2 days. The genomic DNA was isolated according to a modification of the method of Kieser *et al*. (Kieser *et al*., 2000).

The cell pellet was washed with 30 mL of dH_2O, centrifuged (10 min, 2000 rpm) and washed again with 50 mL of 10 mM EDTA pH 8. After centrifugation (10 min, 2000 rpm) the mycelia were resuspended in 12.5 mL of SET buffer (75 mM NaCl, 25 mM EDTA pH 8, 20 mM Tris-HCl pH 7.5) containing 250 µL of lysosyme (50 mg/mL). The cell suspension was incubated for 3h at 37 ^oC and 350 µL of proteinase K solution (20 mg/mL) and 1.5 mL of 10% SDS was added. The mixture was incubated for a further 2 h at 55 ^oC and 5 mL of 5 M NaCl was added. Then the mixture was mixed and gently homogenised. After cooling to 37 ^oC, 12.5 mL of chloroform was added, mixed by inversion for 30 min at room temperature and centrifuged (15 min, 4000 rpm). The supernatant was transferred to a fresh tube. To precipitate the DNA, 0.6 volumes of ice-cold isopropanol was added. The precipitated DNA was washed with 70% EtOH. The DNA was then resuspended in TE buffer or autoclaved dH_2O and was stored at 4 ^oC.

The genome of *S. longisporusruber* was sequenced by the Earlham Institute using a PacBio RS II sequencer (2 SMRT cells) and the sequencing reads were assembled into a single linear contiguous sequence by Dr Emmanuel Lorenzo de los Santos (University of Warwick).

6.2.12.2 Plasmid or cosmid isolation from *E. coli*

Plasmid or cosmid were isolated from *E. coli* using the method described by Sambrook and Russell (2001).

An overnight liquid culture (5 mL of LB liquid medium containing appropriate antibiotic(s) was prepared. After centrifugation, the pellet was resuspended in 100 µL of cold Solution I (25 mM Tris-HCl pH 8, 10 mM EDTA pH 8, 50 mM Glucose). Cells were lysed by adding 200 µL of Solution II (0.2 N NaOH, 1% SDS) and the mixture was gently mixed (lysis was carried out for a maximum of 5 min). 150 µL of cold Solution III (5 M Potassium acetate 60 mL, Glacial acetic acid 11.5 mL, H2O 28.5 mL) was added to precipitate proteins. The solution was mixed by inverting the tube by hand 10 times and the tube was centrifuged (10 min, 14000 rpm, 4 °C). The supernatant was transferred to a new clean tube. 200 µL of phenol-chloroform-isoamyl alcohol (25:24:1) was added was vortexed for 30 s. The mixture was centrifuged (5 min, 14000 rpm) and the water layer was transferred to a clean tube. The plasmid DNA was precipitated with 500 µL of ice-cold isopropanol, the DNA was precipitated on ice for 10 min and the sample was centrifuged (10 min, 14000 rpm, 4 ^{o}C). The DNA was washed with 70% EtOH, centrifuged and dried. DNA was resuspended in TE buffer or autoclaved dH_2O.

6.2.13 Growth of *Streptomyces* and extraction of prodiginines

To analyse prodiginine production, solid medium was used. R5 agar plates were overlaid with sterile semi-permeable membranes. 10 µl of a spore suspension of each mutant were spread separately on a plate to which 100 µl of sterile water had been added. After 5 to 7 days of incubation at 30 °C, mycelia were scraped off and collected in separate tubes. Prodiginines were extracted with ~ 5 ml/plate of mixture of MeOH

acidified (to pH 3.0) by HCl. After vortex and sonication (2 x 20 s), the extracts were centrifuged and the supernatant were collected for analysis. For large scale extraction of undecylprodigiosin and metacycloprodigiosin, 10 plates from each were used. Following the extraction with the solvent, the compound was dried over $MgSO_4$ and concentrated under vacuum. The residue was purified by HPLC to yield the product.

6.2.14 Feeding and expressing *mcpH, mcpHG* and *rphG* genes

For feeding experiment, after 2 days of incubating 10 µL of the *Streptomyces* spore suspension on R5 agar plate, 2-3 mg of each synthetic 2-UP (or 11-methyldodocylpyrrole) and MBC were dissolved in 100 µl of MeOH and DMSO respectively. Both were mixed and were added to tha plate as droplets. Incubation was continued for a further 3 to 5 days, the biomass was scrapped off into a glass vial and the extraction was carried out as described below (section 6.2.15.2).

6.2.15 HPLC and LC-MS experimental

6.2.15.1 HPLC purification

Prodiginines were purified on an Agilent 1200 HPLC instrument fitted with an Agilent Zorbax XDB-C18 column (21.2 x 150 mm, 5 µm, 25°C) with a flow rate of 20 ml / min. Samples were monitored at 533 nm and collected using an automated fraction collector set to a threshold of 400 mAU.

HPLC grade formic acid (1 ml / l) was added to HPLC grade methanol and water and filtered through Nylon 0.2 µm membranes (GE Healthcare) immediately prior to use (Table 6.7).

Table 6.7 HPLC conditions used to purify metacycloprodigiosin and undecylprodigiosin.

Time (mins)	Water % (0.1% formic acid	MeOH % (0.1% formic acid
0.00	60	40
5.00	60	40
20.00	0	100
25.00	0	100
30.00	60	40

6.2.15.2 LC-MS

For LC-MS analysis, the *streptomyces* strain was grown on an appropriate medium. After 5 to 7 days of incubation at 30 °C, mycelia were scraped off and extracted with 5 mL/plate of MeOH (acidified with HCl to 0.1 M). After vortexing and sonication (2 x 20s) the extract was centrifuged and 100 ul of the supernatant was daliuted with 0.4 ml methanol. The suspension was filtered through a 0.2 um filter. The volume was cpmpleted with methanol to 1 ml and analysed by LC-MS.

To analyse culture extracts/culture supernatants of *Streptomyces* strains, Liquid chromatography – Mass Spectrometry (LC-MS) was performed on a Dionex Ultimate 3000 HPLC instrument fitted with an Agilent Zorbax Eclipse Plus C18 600 Bar (2.1 x 100 mm, 1.8 µm, 25°C)connected to a Bruker MaXis Impact mass spectrometer [ESI in positive ion mode; full scan 50-2500 *m/z*; end plate offset, -500 V; capillary, -4500 V; nebulizer gas (N_2), 1.4 bar; dry gas (N_2), 8 L/min; dry temperature, 200 °C] (Table 6.8).

Table 6.8 Gradient elution profile used in LC-MS analysis of prodiginine production

Time (mins)	Water % (0.1% formic acid	MeOH % (0.1% formic acid
0	95	5
5	95	5
17	0	100
23	0	100
25	95	5

References

7. References

Alberts, A.; Vagelos, P. R. *The Enzymes*, **1972** *6*, 37-82.

Ansari, M. Z.; Yadav, G.; Gokhale, R. S.; Mohanty, D. *Nucleic Acids Res*. **2004**, *32*, 404-413.

Aoki, Y.; Matsumoto, D.; Kawaide, H.; Natsume, M. *The Journal of Antibiotics*. **2011**, *64*, 607-611.

Axcell, B. C.; Geary, P. J. *Biochem. J.* **1975**, *146*, 173-183.

Azuma, T.; Watanabe, N.; Yagisawa, H.; Hirata, H.; Iwamura, M.; Kobayashi, Y. *Immunopharmacology* **2000**, *46*, 29-37.

Barona-Gomez, F.; Lautru, S.; Francou, F. X.; Leblond, P.; Pernodet, J. L.; Challis, G.L. *Microbiology* **2006**, *152*, 3355-3366.

Barona-Gómez, F.; Wong, U.; Giannakopulos A. E.; Derrick, P. J ;Challis, G. L. *J. Am. Chem. Soc.*, **2004**, 126, 16282-16283.

Barry, S. M.; Challis G. L. *ACS Catal* **2013**, *3*, 2362–2370.

Bennett, J. W.; Bentley, R. *Adv*. *Appl*. *Microbiol*. **2000**, *47*, 1-32.

Bentley, S. D.; Brown, S.; Murphy, L. D.; Harris, D. E.; Quail, M. A.; Parkhill, J.; Barrell, B. G.; McCormick, J. R.; Santamaria, R. I.; Losick, R.; Yamasaki, M.; Kinashi, H.; Chen, C. W.; Chandra, G.; Jakimowicz, D.; Kieser, H. M.; Kieser, T.; Chater, K. F. *Mol*. *Microbiol*. **2004**, *51*, 1615-1628.

Bentley, S. D.; Chater, K. F.; Cerdeno-Tarraga, A. M.; Challis, G. L.; Thomson, N. R.; James, K. D.; Harris, D. E.; Quail, M. A.; Kieser, H.; Harper, D.; Bateman,

A.; Brown, S.; Chandra, G.; Chen, C. W.; Collins, M.; Cronin, A.; Fraser, A.; Goble, A.; Hidalgo, Hornsby, T.; Howarth, S.; Huang, C. H.; Kieser, T.; Larke, L.; Murphy, L.; Oliver, O'Neil, S.; Rabbinowitsch, E.; Rajandream, M. A.; Rutherford, K.; Rutter, S.; Seeger, K.; Saunders, D.; Sharp, S.; Squares, R.; Squares, S.; Taylor, K.; Warren, T.; Wietzorrek, A.; Woodward, J.; Barrell, B. G.; Parkhill, J.; Hopwood, D. A. *Nature* **2002**, *417*, 141-147.

Bibb, M. J. *Curr*. *Opin*. *Microbiol*. **2005,** *8*, 208-215.

Black, P.N.; DiRusso, C. C. *Biochim*. *Biophys*. *Acta*. **1994**, *27*, 123–145.

Boonlarppradab C.; Kauffman C.; Jensen; Fenical W. *Org*. *Lett*. **2008**, *10*, 5508.

Brockmann, H.; Pini, H.; Plotho, O. *Chem*. *Ber*. **1950**, *83*, 161.

Buchholz, T.J; Rath, M. C.; Lopanik, N. B.; Gardner, N. P.; Hakansson, K.; Sherman, D. H. *Chemistry and Biology*, **2010**, *17*, 1092-1100.

Bugg T. D.; Ramaswamy S, *Curr*. *Opin*. *Chem*. *Biol*. **2008,** *12*, 134-140.

Cerdeño, A. M.; Bibb, M. J.; Challis, G. L. *Chem*. *Biol*. **2001**, *8*, 817-829.

Challis, G. L. *J*. *Ind*. *Microbiol*. *Biotechnol*. **2014**, *41*, 219-232.

Challis, G. L.; Ravel, J. *FEMS Microbiol*. *Lett*. **2000**, *187*, 111-114.

Challis, G. L.; Ravel, J.; Townsend, C. A. *Chem*. *Biol*. **2000**, *7*, 211-224.

Corre, C.; Challis, G. L. *ChemBioChem*. **2005**, *6*, 2166-2170.

Corre, C.; Song, L.; O'Rourke, S.; Chater, K. F.; Challis, G. L. *Proc Natl*. *Acad*. *Sci*. *U.S.A*. **2008,** *105*,17510-17515.

Cortes, J.; Wiesmann, K. E.; Roberts, G. A.; Brown, M. J.; Staunton, J.; Leadlay, P. F. *Science* **1995**, *268*, 1487-1489.

Danial, N. N.; Korsmeyer, S. J. *Cell* **2004**, *116*, 205-219.

Del Vecchio, F.; Petkovic, H.; Kendrew, S. G.; Low. L. *J. Ind. Microbiol. Biotechnol.* **2003**, *30*, 489-494.

Caffrey, P. *Chem. Biol.* **2003**, *4*, 654-657.

Dornadio, S.; Monciardini, P.; Sosio, M. *Nat. Prod. Rep*. **2007**, *24*, 1073.

Du, Y.; Wang, Y.; Huang, T.; Tao1, M.; Deng, Z.; Lin, S. *BMC Microbiol.* **2014**, https://doi.org/10.1186/1471-2180-14-30.

Eaton, R. W.; Chapman, P. J. J. *J Bacteriol* **1992**, *174*, 7542-7554.

Elson, S. W.; Baggaley, K. H.; Gillett, J.; Holland, S.; Nicholson, N. H.; Sime, J. T.;Wroniecki, S. R. *J. Chem. Soc. Chem. Commun*. **1987**, *22*, 1736-1738.

Fernandez-Moreno, M. A.; Martinez, E.; Boto, L.; Hopwood, D. A.; Malpartida, F. *J. Biol. Chem*. **1992,** *267*, 19278-19290.

Ferraro DJ; Gakhar L.; Ramaswamy S. *Biochem. Biophys. Res Commun*. **2005**, *338*, 175-190.

Finking, R.; Marahiel, M. A. *Annu. Rev. Microbiol*. **2004**, *58*, 453-488.

Foerstner, K. U.; Doerks, T.; Creevey, C. J.; Doerks, A.; Bork, P. *PLoS One* **2008**, *3*, 3515.

Furstner, A. *Angew. Chem. Int Ed Engl* **2003**, *42*, 3582-3603.

Garwin, J. L.; Klages, A. L.; Cronan, J. E. Jr. *J Biol*. *Chem*. **1980**, *255*, 3263-3265.

Gaughran, E. R. L. *Trans*. *N. Y. Acad. Sci*. **1969**, *31*, 3-24.

Gerber, N. N. *CRC Crit*. *Rev*. *Microbiol*. **1975**, *3*, 469-485.

Gerber, N. N. *J Antibiot (Tokyo)* **1975**, *28*, 194-199.

Gerber, N.; McInnes, A.; Smith, D.; Walter, J.; Wright, J.; Vining, L. *Can*. *J*. *Chem*. **1978**, *56*, 1155-1163.

Gibson, D. T.; Koch, J. R.; Kallio, R. E. *Biochemistry*, **1968**, *7*, 2653-2662.

Giessen, T.W.; Franke, K. B.; Knappe, T. A.; Kraas, F. I.; Bosello M.; Xie X.; Linne, U.; Marahiel, M. A. *J. Nat. Prod.* **2012**, 75, 905-914.

Gómez-Escribano, J. P.; Song, L.; Fox, D. J.; Yeo, V.; Bibb, M. J.; and Challis, G. L. *Chem*. *Sci*. **2012**, *3*, 2716-2720.

Guchhait R. B.; Polakis S. E.; Dimroth P.; Stoll E.; Moss J.; Lane M. D. *J*. *Biol*. *Chem*. **1974**, *249*, 6633–6645.

Gust, B.; Challis, G. L.; Fowler, K.; Kieser, T.; Chater, K. F. *Proc Natl*. *Acad*. *Sci*. *U.S.A*. **2003**, *100*, 1541-1546.

Hammerschmidt, F. *J. Chem. Soc. Perkin*. *Trans*. **1991**, *1*, 1993-1996.

Haneishi, T.; Terahara, A.; Arai, M.; Hata, T.; Tamura, C. *J*. *Antibiot. (Tokyo)* **1974**, *27*, 393-399.

Harris, A. K.; `liamson, N. R.; Slater, H.; Cox, A.; Abbasi, S.; Foulds, I.; Simonsen, H. T.; Leeper, F. J.; Salmond, G. P. *Microbiology* **2004**, *150*, 3547-3560.

Haug, I.; Weissenborn, A.; Brolle, D.; Bentley, S.; Kieser, T.; Altenbuchner, J. *Microbiology* **2003,** *149*, 505–513.

Hayakawa, Y.; Nagatsuka, S. Y.; Kawasaki, T. *J Antibiot (Tokyo)* **2009**, *62*, 531-532.

Haydock, S. F.; Aparicio, J. F.; Molnár, I.; Schwecke, T.; Khaw, L. E.; Ariane König, A.; Marsden A. F. A.; Galloway, I. S.; Staunton, J.; Peter F. Leadlay, *FEBS Lett.* **1995**, *374*, 246-248.

Haynes, S. W.; Sydor, P. K.; Stanley, A. E.; Song, L.; Challis, G. L. *Chem. Commun*. **2008**, 1865-1867.

Higgins, L. J.; Yan, F.; Liu, P.; Liu, H. W.; Drennan, C. L. *Nature* **2005**, *437*, 838-44.

Hojati, Z.; Milne, C.; Harvey, B.; Gordon, L.; Borg, M.; Flett, F.; Wilkinson, B.; Sidebottom, P. J.; Rudd, B. A.; Hayes, M. A.; Smith, C. P.; Micklefield, J. *Chem Biol*. **2002**, *9*, 1175-1187.

Hopwood, D. A. *Chem Rev*. **1997**, *97*, 2465-2498.

Hornemann, U.; Hopwood, D.A., H. *Tetrahedron Lett.* **1978**, *33*, 2977-2978.

Imbert, M.; Bechet, M.; Blondeau, R. *Curr. Microbiol*. **1995**, *31*, 129-133.

Jiang H.; Parales R. E.; Lynch N. A.; Gibson D. T. *J. Bacteriol*. **1996**, *178*, 3133-3139

Jiang, J.; He, X.; Cane, D.E. *J. Am. Chem. Soc.*, **2006**, 128 (25), 8128–8129

Jung, D.; Rozek, A.; Okon, M.; Hancock, R. E. *Chem Biol*. **2004**, *11*, 949-957.

Karl Poralla, K.; Muth, G.; Härtner, T.; *FEMS Microbiol. Lett.* **2000**, *189*, 93-95.

Kauppi, B.; Lee, E.; Carredano, E.; Parales, R. E.; Gibson, D. T.; Eklund, H.; Ramaswamy, S. *Structure* **1998**, *6*, 571-586.

Kawasaki, T.; Sakurai, F.; Hayakawa, Y. *J Nat Prod* **2008**, *71*, 1265-1267.

Kawasaki, T.; Sakurai, F.; Nagatsuka, S. Y.; Hayakawa, Y. *J Antibiot (Tokyo)* **2009**, *62*, 271-276.

Keatinge-Clay, A. T. *Chemistry & Biology*, **2007**, *14*, 898-908.

Keatinge-Clay, A. T. *J. Mol. Biol.* **2008**, *384*, 941-953.

Kempter, C.; Kaiser, D.; Haag, S.; Nicholson, G.; Gnau, V.; Walk, T.; Gierling, K. H.; Decker, H.; Zahner, H.; Jung, G.; Metzer, J. W. *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 498-501.

Kieser, T.; Bibb, M. J.; Buttner, M. J.; Chater, K. F.; Hopwood, D. A. Practical Practical Strepto-myces Genetics. 2 ed.; The John Innes Centre Fundation: Norwich, **2000**.

Kimata, S; Izawa, M; Kawasaki, T; Hayakawa, Y. *J Antibiot*. **2016**, *70*, 196-199.

Kirby, R.; Hopwood, D. A. *J. Gen. Microbiol.* **1977**, *98*, 239-252.

Komatsu, M., Tsuda, M., Omura, S., Oikawa, H., Ikeda, H.,. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 7422-7427.

Konomi, T.; Herchen, S.; Baldwin, J. E.; Yoshida, M.; Hunt, N. H.; Demain, A. L. *Biochem. J.* **1979**, *184*, 427-430.

Kwan D. H.; Sun. Y.; Schulz, F.; Hong H.; Popovic B.; Sim-Stark , J.C. Haydock, S. F.; Leadlay, P. F. *Chem. Biol.* **2008**, *15*, 1231-1240.

Lakey, J. H.; Lea, E. J.; Rudd, B. A.; Wright, H. M.; Hopwood, D. A. *J. Gen. Microbiol*. **1983**, *129*, 3565-3573.

Lautru, S.; Deeth, R. J.; Bailey, L. M.; Challis, G. L. *Nat. Chem. Biol*. **2005**, *1*, 265-269.

Lee, J. K; Simurdiak, M; Zhao, H. *J. Biol. Chem.* **2005**, *280*, 36719-36727.

Lee, M. H.; Kataoka, T.; Honjo, N.; Magae, J.; Nagai, K. *Immunology* **2000**, *99*, 243-248.

Lei, L.; Waterman, M. R.; Fulco, A. J.; Kelly, S. L.; Lamb, D. C. *Proc Natl. Acad. Sci. U.S.A.* **2004**, *101*, 494-499.

Liu, P.; Murakami, K.; Seki, T.; He, X.; Yeung, S. M.; Kuzuyama, T.; Seto, H.; Liu, H. *J. Am. Chem. Soc*. **2001**, *123*, 4619-4620.

Madigan, M.; Martinko, J. Brock Biology of Microorganisms. 11 ed.; Prentice Hall: New Jersey, **2005**.

Magae, J.; Miller, M. W.; Nagai, K.; Shearer, G. M. *J Antibiot (Tokyo)* **1996**, *49*, 8690.

Malpartida, F.; Niemi, J.; Navarrete, R.; Hopwood, D. A. *Gene* **1990**, *93*, 91-99.

Marini, P.; Li, S. J.; Gardiol, D.; Cronan, J. E. Jr.; de Mendoza, D. *J. Bacteriol*. **1995**, *177*, 7003–7006.

McAlpine, J. B.; Bachmann, B. O.; Piraee, M.; Tremblay, S.; Alarco, A.M; Zazopoulos, E.; Chris M. Farnet, CM. *J. Nat. Prod.* **2005**, *68*, 493-496.

McGuire, J. M.; Binch, R. L.; Anderson, R. C.; Boaz, H. E.; Flynn, E. H.; Powell, M.; Smith, J. W. *Antibiot. Chemother*. **1952**, *2*, 281-283.

McMurry, J.; Begley, T. The Organic Chemistry of Biological Pathways. Roberts and Company Publisher: Englewood CO, **2005**.

Mo, S.; Kim, B. S.; Reynolds, K. A. *Chem*. *Biol*. **2005**, *12*, 191-200.

Mo, S.; Sydor, P. K.; Corre, C.; Alhamadsheh, M. M.; Stanley, A. E.; Haynes, S. W.; Song, L.; Reynolds, K. A.; Challis, G. L. *Chem*. *Biol*. **2008**, *15*, 137-148.

Mortellaro, A.; Songia, S.; Gnocchi, P.; Ferrari, M.; Fornasiero, C.; D'Alessio, R.; Isetta, A.; Colotta, F.; Golay, J. *J*. *Immunol*. **1999**, *162*, 7102-7109.

Nguyen, M.; Cencic, R.; Ertel F; Bernier, C.; Pelletier J.; Roulston, A.; Silvius J. R.; Shore, G. C. *BMC Cancer* **2015**, *15*, 568.

Nguyen, M.; Marcellus, R. C.; Roulston, A.; Watson, M.; Serfass, L.; Murthy Madiraju, S. R.; Goulet, D.; Viallet, J.; Belec, L.; Billot, X.; Acoca, S.; Purisima, E.; Wiegmans, A.; Cluse, L.; Johnstone, R. W.; Beauparlant, P.; Shore, G. C. *Proc*. *Natl*. *Acad*. *Sci*. *U*.*S*.*A*. **2007**, *104*, 19512-19517.

Ohta T.; Chakrabarty S.; Lipscomb JD.; Solomon EI., *J. Am. Chem. Soc*. **2008**, *130*, 1601-1610.

Papireddy K.; Smilkstein M.; Kelly J.X.; Salem S.M.; Alhamadsheh M.; Haynes S.W.; Challis G.L.; Reynolds K.A. *J*. *Med*. *Chem*. **2011**, *54*, 5296-5306.

Pastor, J. M.; Salvador, M.; Argandoña, M.; Bernal, V.; Reina-Bueno, M.; Csonka, L. N.; Iborraa, J. L.; Vargasb, C.; Nietob, J. J.; Cánovasa, M. *Biotechnol*. *Adv*. **2010**, *28*, 782-801.

Raynal, A.; Friedmann, A.; Tuphile, K.; Guerineau, M.; Pernodet, J. L. *Microbiology* **2002,** *148* (Pt 1), 61-67.

Rebets, Y; Brötz, E; Manderscheid, N; Tokovenko, B; Myronovskyi, M; Metz, P; Petzke, L; Luzhetskyy, A; *Angew*. *Chem*. *Int*. *Ed*. *Engl*. **2015**, *54*, 2280-2284.

Reshetnikov, A. S.; Khmelenina, V. N.; Mustakhimov, I. I.; Kalyuzhnaya, M.; Lidstrom, M.; Trotsenko, Y.A. . *Extremophiles* **2011**, ***15***, 653-663.

Roach, P. L.; Clifton, I. J.; Fulop, V.; Harlos, K.; Barton, G. J.; Hajdu, J.; Andersson, I.; Schofield, C. J.; Baldwin, J. E. *Nature* **1995**, *375*, 700-704.

Rudd, B. A.; Hopwood, D. A. *J*. *Gen*. *Microbiol*. **1980**, *119*, 333-340.

Rutledge, P. J., and Challis, G. L. **2015**, *Nat*. *Rev*. *Microbiol*. 13, 509-523.

Sambrook, J.; Russell, D. W. Molecular Cloning - A Laboratory Manual. 3 ed.; Cold Spring Harbor Laboratory Press: New York, **2001**.

Saxena, P.; Yadav, G.; Mohanty, D.; Gokhale, R. S. *J*. *Biol*. *Chem*. **2003**, *278*,44780-44790.

Seto, H.; Hidaka, T.; Kuzuyama, T.; Shibahara, S.; Usui, T.; Sakanaka, O.; Imai, S. *J*. *Antibiot*. **1991**, *44* (1286).

Song, L.; Barona-Gomez, F.; Corre, C.; Xiang, L.; Udwary, D. W.; Austin, M. B.; Noel, J. P.; Moore, B. S.; Challis, G. L. *J*. *Am*. *Chem*. *Soc*. **2006**, *128*, 14754-14755.

Song, L.; Barona-Gomez, F.; Corre, C.; Xiang, L.; Udwary, D. W.; Austin, M. B.; Noel, J. P.; Moore, B. S.; Challis, G. L. *J*. *Am*. *Chem*. *Soc*. **2006**, *128*, 14754-14755.

Stanley, A. E.; Walton, L. J.; Kourdi Zerikly, M.; Corre, C.; Challis, G. L. *Chem. Commun*. **2006**, 3981-3983.

Staunton, J.; Wilkinson, B. *Chem. Rev*. **1997**, *97*, 2611-2630.

Strohl, W. R. *Nucleic Acids Res* **1992**, *20*, 961-974.

Streptomyces Genetics. 2 ed.; The John Innes Centre Fundation: Norwich, **2000**.

Summers, R. M; Louie, T. M; Yu C. L; Gakhar L; Louie K. C.; Subramanian M. *J. Bacteriol*. **2012**, *194*, 2041-2049.

Sydor, P. K. S. Elucidation of the Prodiginine Biosynthetic Pathway in *Streptomyces coelicolor* A3(2). PhD Thesis, University of Warwick, Coventry, **2010**.

Sydor, P.K.;Barry S.M.; Odulate O. M.;Barona-Gomez F.;Haynes S. W.;Corre C.; Song L.; Challis G. L.; *Nat. Chem*. **2011**, *3*, 388-392.

Takano, E. *Curr. Opin. Microbiol*. **2006**, *9*, 287-294.

Takano, E.; Kinoshita, H.; Mersinias, V.; Bucca, G.; Hotchkiss, G.; Nihira, T.; Smith, C. P.; Bibb, M.; Wohlleben, W.; Chater, K. *Mol. Microbiol*. **2005**, *56*, 465-479.

Takano, E.; Nihira, T.; Hara, Y.; Jones, J. J.; Gershater, C. J.; Yamada, Y.; Bibb, M. *J. Biol. Chem*. **2000**, *275*, 11010-1106.

Thomas, M. G.; Burkart, M. D.; Walsh, C. T. Chem. Biol. 2002, 9, 171-184.

Traxler, M. F.; Watrous, J. D.; Alexandrov, T.; Dorrestein P. C.; Kolter, R. *mBio*. **2013**, *4*, e00459-13.

Trudel, S.; Li, Z. H.; Rauw, J.; Tiedemann, R. E.; Wen, X. Y.; Stewart, A. K. *Blood* **2007**, *109*, 5430-5438.

Tsuji RF.; Yamakoshi, J.; Uramoto, M.; Koshino H.; Saito M.; Kikuchi M.; Masuda T. *Immunopharmacology* **1995**, *29*, 79-87.

Tsuji, R.F.; Magae, J.; Yamashita, M.; Nagai, K.; Yamasaki, M. *J Antibiot* (Tokyo) **1992**, *45*, 1295-1302.

Umezawa, K.; Nakazawa, K.; Ikeda, Y.; Naganawa, H.; Kondo, S. *J. Org. Chem.* **1999**, *64*, 3034-3038.

Unpublished data from Challis group.

Van der Geize, R.; Yam, K.; Heuser, T.; Wilbrink, M. H.; Hara, H.; Anderton, M. C.; Sim, E.; Dijkhuizen, L.; Davies, J. E.; Mohn, W. W.; Eltis, L. D. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1947-1952.

Varghese F. S.; Rausalu K.; Hakanen M.; Saul S.; Kümmerer, B. M; Susi, P.; Merits, A.; Ahola, T. *Antimicrob. Agents Chemother.* **2017**, *61*, e02227-16.

Vertesy, L.; Ehlers, E.; Kogler, H.; Kurz, M.; Meiwes, J.; Seibert, G.; Vogel, M.; Hammann, P. *J. Antibiot. (Tokyo)* **2000**, *53*, 816-827.

Wandersman, C.; Delepelaire, P. *Annu. Rev. Microbiol.* **2004**, *58*, 611-647.

Wasserman, H. H.; Keith, D. D.; Nadelson, J. *J. Am. Chem. Soc.* **1969**, *91*, 1264-1265.

Wasserman, H. H.; Rodgers, G. C.; Keith, D. D. *J Am Chem. Soc.* **1969**, *91*, 1263-1264.

Wasserman, H. H.; Shaw, C. K.; Sykes, R. J. *Tetrahedron Lett.* **1974**, *15*, 2787-2790.

Wasserman, H. H.; Skles, R. J.; Peverada, P.; Shaw, C. K.; Cushley, R. J.; Lipsky, C. R. *J. Am. Chem. Soc.* **1973**, *95*, 6874-6875.

Weber, J. M.; Leung, J. O.; Maine, G. T.; Potenz, R. H.; Paulus, T. J.; DeWitt, J. P. *J. Bacteriol.* **1990**, *172*, 2372-2383.

Weber, T.; Blin, K.; Duddela, S.; Krug, D.; Uk Kim, H.; Robert Bruccoleri, R.; Yup Lee, S.; Fischbach, A. M.; Müller, R.; Wohlleben, W.; Breitling, R.; Takano, E.; Medema, M. H. *Nucleic Acids Res.* **2015**, *43*, 237-243.

Whicher J.R.; Florova G.; Sydor P. K.;Singh R.; Alhamadsheh M.; Challis G.L.; Reynolds K. A.; Smith J.L. *J. Biol. Chem.* **2011**, *286*, 22558-22569.

Williamson, N. R.; Fineran, P. C.; Gristwood, T.; Chawrai, S. R.; Leeper, F. J.; Salmond, G. P. *Future Microbiol* **2007**, *2*, 605-618.

Williamson, N. R.; Simonsen, H. T.; Ahmed, R. A.; Goldet, G.; Slater, H.; Woodley, L.; Leeper, F. J.; Salmond, G. P. *Mol. Microbiol.* **2005**, *56*, 971-989.

Williamson, Neil R; Fineran, Peter C; Leeper, Finian J; Salmond, George P. C. *Nat. Rev. Microbiol.* **2006**, *4*, 887-899.

Withall, D M.; Haynes S W; Challis Gregory L. *J. Am. Chem. Soc.*, **2015**, *137*, 7889-7897.

Wright, L. F.; Hopwood, D. A. *J. Gen. Microbiol.* **1976**, *96*, 289-297.

Yamanaka, K; Reynolds, K. A; Kersten, R. D; Ryan, K. S; Gonzalez, D. J; Nizet, V; Dorrestein, P. C; Moore, B. S. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, 111:1957-1962. [PubMed: 24449899]

Zerbe, K.; Woithe, K.; Li, D. B.; Vitali, F.; Bigler, L.; Robinson, J. A. *Angew Chem Int*. *Ed*. *Engl*. **2004**, *43*, 6709-6713.

Zhang, J.; Shen, Y.; Liu, J.; Wei, D. *Biochem Pharmacol* **2005**, *69*, 407-414.

Zhang, Z.; Ren, J.; Stammers, D. K.; Baldwin, J. E.; Harlos, K.; Schofield, C. J. *Nat*. *Struct*. *Biol*. **2000**, *7*, 127-133.

Zhao, B.; Lin, X.; Lei, L.; Lamb, D. C.; Kelly, S. L.; Waterman, M. R.; Cane, D. E. *J*. *Biol*. *chem*. **2008**, *283*, 8183-8189.

Zhu. D; Liu. J; Han. R; Shen. G; Long. Q; Wei. X; Liu, D, *J*. *Microbiol*. **2014**, *52*, 139-147.

Appendix

Gene clusters from *Streptomyces longispororuber* blasted against NCBI.

(ID - identity, a. a. - amino acid, Hypo - hypothesised protein, Low - low identity)

Gene No.	Gene	Organism	ID	a.a.	Accession No.
	Cluster 1 Melanin				
32	DUF3344 domain-containing protein	*Streptomyces* sp. 3124.6	57%	382	WP_079662163.1
37	membrane protein	*S. aureocirculatus*	80%	224	WP_030559644.1
38	tyrosinase	*Streptomyces* sp. NRRL S-813	58%	129	WP_078917081.1
39	tyrosinase	*S. aurantiacus*	98%	289	WP_037660750.1
40	chaplin	*S. aurantiacus*	99%	87	EPH42772.1
42	peptidase	*S. aureocirculatus*	82%	250	WP_078964781.1
43	chemotaxis protein CheB	*S. aurantiacus*	98%	207	WP_016642290.1
44	chemotaxis protein CheR	*Streptomyces* sp. NRRL B-1347	81%	621	WP_030683960.1
	Cluster 2 Other KS				
37	dTMP kinase	*S. aurantiacus*	95%	1090	WP_016644694.1
338	DNA polymerase III subunit delta	*S. aurantiacus*	99%	401	WP_016644695.1
339	alpha/beta hydrolase		99%	484	WP_016644696.1
340					
341	chaplin	*S. aurantiacus*	96%	79	WP_037663470.1
342	StuI	*S. tubercidicus*	96%	256	ADZ31412.1
343	SAM-dependent methyltransferase	*S. albolongus*	86%	699	WP_084749176.1
344	cupin	*S. phaeochromogenes*	90%	122	WP_055610342.1
345					
346	acyl-CoA synthetase	*S. ruber*	82%	525	WP_078861769.1
347	TetR/AcrR family transcriptional regulator	*S. aurantiacus*	96%	234	WP_016644748.1
348					
349	putative Bifunctional epoxide hydrolase 2	*S. aurantiacus* JA 4570	97%	329	EPH40368.1
350					
351	RNA polymerase subunit sigma	*Streptomyces* sp. NRRL S-920	83%	183	WP_078594686.1
352					
353	damage-inducible protein DinB	*S. aurantiacus*	92%	180	WP_078621687.1
354					
355	transcriptional regulator	*S. aurantiacus*	86%	275	WP_016644755.1
356	DUF397 domain-containing protein	*S. aurantiacus*	68%	94	WP_078621685.1
357					
358	3-oxoacyl-ACP synthase	*Nocardiopsis trehalosi*	51%	349	WP_084469302.1
359	NADP-dependent 3-hydroxy acid dehydrogenase YdfG	*Glycomyces sambucus*	59%	252	SDL67379.1
360	DNA-binding response regulator	*S. aurantiacus*	98%	214	WP_016644757.1
361	sensor histidine kinase	*S. ruber*	82%	424	WP_030359722.1
362	3-oxoadipate enol-lactonase	*S. aurantiacus*	92%	270	WP_016644759.1
363	putative Fluoroacetate dehalogenase Pimeloyl-ACP methyl ester carboxylesterase	*S. aurantiacus* JA 4570 *Saccharopolyspora shandongensis*	99% 81%	307	EPH40378.1 SDX16300.1
364	ATP-binding protein	*S. aurantiacus*	93%	1159	WP_052029195.1
365					
366	AfsR family transcriptional regulator	*S. silvensis*	84%	937	WP_058850266.1
367	putative Sensory transduction protein	*S. aurantiacus* JA 4570	99%	268	EPH42364.1

	regX3				
368	two-component sensor histidine kinase	*S. aurantiacus*	97%	707	WP_052029196.1
369	AfsR family transcriptional regulator	*S. bottropensis*	62%	1018	WP_028799103.1
370	bifunctional 3'-5' exonuclease/DNA polymerase	*S. aurantiacus*	97%	567	WP_016645570.1
	Cluster 3 T2 PKS				
483	redox-sensing transcriptional repressor Rex	*S. aurantiacus*	100%	216	WP_016640256.1
484	glutamyl-tRNA reductase	*S. aurantiacus*	99%	554	WP_078621124.1
485	hydroxymethylbilane synthase	*S. aurantiacus*	99%	324	WP_016640254.1
486	bifunctional uroporphyrinogen-III C-methyltransferase/uroporphyrinogen-III synthase	*S. aurantiacus*	99%	465	WP_016640253.1
487	delta-aminolevulinic acid dehydratase	*S. aurantiacus*	99%	329	WP_052029033.1
488	NADPH quinone oxidoreductase	*S. aureocirculatus*	83%	334	WP_030565188.1
489	6-phosphogluconate dehydrogenase	*S. aureocirculatus*	81%	296	WP_030565190.1
490					
491					
492	transcriptional regulator	*S. aureocirculatus*	68%	503	WP_030565196.1
493	PLP-dependent aminotransferase family protein	*S. aurantiacus*	99%	415	WP_016640246.1
494	arginine--tRNA ligase	*S. aurantiacus*	99%	592	WP_016640244.1
495	lysine--tRNA ligase	*S. aurantiacus*	99%	580	WP_016640243.1
496					
497					
498	alpha/beta hydrolase	*S. aurantiacus*	98%	304	WP_016640240.1
499	polyketide synthase	*S. aurantiacus*	98%	319	WP_016640239.1
500	monensin polyketide synthase ketoacyl reductase	*S. aurantiacus*	99%	261	WP_016640238.1
501	polyketide synthase acyl carrier protein actinorhodin polyketide synthase	*S. aurantiacus* *Streptomyces* sp. NRRL F-2580	97% 85%	88	WP_016640237.1 WP_030718886.1
502	putative Actinorhodin polyketide beta-ketoacyl synthase 2	*S. aurantiacus* JA 4570	99%	398	EPH44817.1
503	beta-ACP synthase	*S. aurantiacus*	99%	423	WP_037658548.1
504	polyketide synthase	*Streptomyces* sp. NRRL F-5755	89%	216	KOT87189.1
505	nuclear transport factor 2 family protein	*S. aurantiacus*	99%	156	WP_016640233.1
506	putative Ferredoxin-1	*S. aurantiacus* JA 4570	100%	64	EPH44813.1
507	cytochrome P450	*S. aurantiacus*	100%	406	WP_052029032.1
508	DNA-binding response regulator	*S. aurantiacus*	97%	269	WP_078621122.1
509	nuclear transport factor 2 family protein	*S. aurantiacus*	99%	142	WP_016640229.1
510	MFS transporter	*S. aurantiacus*	99%	217	WP_016640227.1
511	MFS transporter	*S. aurantiacus*	99%	504	WP_016640227.1
512	TetR family transcriptional regulator	*S. aurantiacus*	98%	202	WP_016640226.1
513	membrane protein	*S. aureocirculatus*	68%	583	WP_030565204.1
514					
515	DUF3558 domain-containing protein	*S. atriruber*	75%	299	WP_055570158.1
516					
517	RtcB family protein	*S. aurantiacus*	98%	397	WP_016640264.1
518					
519	putative oxidoreductase	*S. aurantiacus*	99%	529	WP_016640265.1
520	YnfA family protein *Streptomyces aurantiacus*		98%	111	WP_016640268.1
521	UPF0060 family protein	*S. aurantiacus*	98%	161	WP_016640268.1
522	GNAT family N-acetyltransferase	*S. aurantiacus*	97%	192	WP_078621126.1
523	serine hydrolase	*S. aurantiacus*	99%	294	WP_016640270.1
524					
525					
526	penicillin-binding protein A	*S. aurantiacus*	98%	491	WP_016640273.1
	Cluster 4 Ladderane NRPS				
888	nitrate reductase	*S. aurantiacus*	97%	241	WP_016639799.1

889	succinate dehydrogenase iron-sulfur subunit	*S. aurantiacus*	100%	256	WP_016639798.1
890	succinate dehydrogenase flavoprotein subunit	*S. aurantiacus*	99%	584	WP_016639797.1
891	succinate dehydrogenase	*Streptomyces* sp. NRRL S-920	100%	164	WP_016639796.1
892	succinate dehydrogenase, cytochrome b556 subunit	*S. aurantiacus*	98%	126	WP_078621033.1
893	OHCU decarboxylase	*S. aurantiacus*	98%	167	WP_078621031.1
894	beta-N-acetylglucosaminidase	*S. aurantiacus*	98%	539	WP_037658172.1
895					
896					
897					
898					
899					
900	two-component sensor histidine kinase	*S. aurantiacus*	93%	372	WP_037658144.1
901	DNA-binding response regulator	*S. aurantiacus*	93%	241	WP_016639785.1
902	ABC transporter ATP-binding protein	*S. aurantiacus*	98%	599	WP_016639784.1
903	Taurine dioxygenase, alpha-ketoglutarate-dependent	*Micromonospora peucetia*	69%	313	SCL70749.1
904	putative transcriptional regulator RedD	*S. aurantiacus*	98%	259	WP_016639780.1
905	flavin reductase	*S. aurantiacus*	99%	171	WP_016639779.1
906	transcriptional regulator	*S. aurantiacus*	97%	324	WP_052028994.1
907	non-ribosomal peptide synthetase	*Nitrospira defluvii*	46%	515	WP_013248703.1
908	LysR family transcriptional regulator	*S. aurantiacus*	97%	303	WP_078621210.1
909	ABC transporter permease	*S. aurantiacus*	96%	853	WP_037659424.1
910	Macrolide export ATP-binding/permease MacB	*S. aurantiacus*	96%	264	WP_016641077.1
911	putative Multifunctional cyclase-dehydratase-3-O-methyl transferase TcmN	*S. aurantiacus* JA 4570	98%	340	EPH43986.1
912	putative Daunorubicin/doxorubicin resistance ATP-binding protein DrrA	*S. aurantiacus*	99%	322	WP_016641079.1
913	MbtH family protein	*S. aurantiacus*	94%	71	WP_016641080.1
914	cytochrome P450	*S. aurantiacus*	99%	409	WP_016641081.1
915	daunorubicin ABC transporter permease	*S. aurantiacus*	99%	272	WP_016641082.1
916	putative Dimodular nonribosomal peptide synthase	*S. aurantiacus* JA 4570	98%	3902	EPH43991.1
917	putative Dimodular nonribosomal peptide synthase	*S. aurantiacus* JA 4570	98%	1539	EPH44537.1
918	2,3-diaminopropionate biosynthesis protein SbnA	*S. aurantiacus*	98%	330	WP_016640521.1
919	tryptophan halogenase	*S. aurantiacus*	99%	532	WP_016640522.1
920	cation:proton antiporter	*S. aurantiacus*	97%	435	WP_016640523.1
921	3-oxoacyl-ACP synthase	*Streptomyces* sp. WM6378	96%	377	WP_052029054.1
922	3-oxoacyl-ACP synthase II	*S. aurantiacus*	99%	407	WP_016640526.1
923	putative Acyl carrier protein	*S. aurantiacus* JA 4570	99%	91	EPH44542.1
924	Regulator protein	*S. aurantiacus*	98%	268	WP_016640527.1
925	2,3-diaminopropionate biosynthesis protein SbnB putative Ornithine cyclodeaminase	*S. aurantiacus* *S. aurantiacus* JA 4570	98% 98%	355	WP_037658814.1 EPH44545.1
926	3-deoxy-7-phosphoheptulonate synthase class II	*S. thermolilacinus*	80%	42	WP_023586044.1
927	PLP-dependent aminotransferase family protein	*S. aurantiacus*	97%	466	WP_016640530.1
928	anthranilate phosphoribosyltransferase	*S. aurantiacus*	98%	352	WP_016640529.1
929	RNA polymerase subunit sigma-70	*S. aurantiacus*	100%	223	WP_078621150.1
930					
931					
932	murein L,D-transpeptidase	*S. aurantiacus*	98%	312	WP_063890396.1
933	MarR family transcriptional regulator	*S. aurantiacus*	98%	175	WP_016645383.1
934	Methyltransferase domain-containing protein	*Streptomyces* sp. 3213	73%	304	SEE35603.1
935	Hypo				

936	glycosyltransferase EpsJ	*S. aurantiacus*	99%	326	WP_016645388.1
937					
938	N-acylneuraminate cytidylyltransferase	*S. aurantiacus*	99%	427	WP_016645389.1
939	sialic acid synthase N-acetylneuraminate synthase	*S. aurantiacus* *Streptomyces* sp. NRRL B-1347	99% 97%	313	WP_016645390.1 WP_030668669.1
940					
941	glycosyltransferase, group 1 family protein	*S. ipomoeae*	57%	758	WP_009313929.1
	Cluster 5 Terpene				
1234	3-phosphoshikimate carboxyvinyltransferase	*S. aurantiacus*	98%	446	WP_016645398.1
1235	membrane protein	*S. aureocirculatus*	96%	240	WP_030560644.
1236	DUF159 family protein	*S. aurantiacus*	99%	271	WP_016645396.1
1237	hydrolase	*S. silaceus*	86%	226	WP_055700426.1
1238	RNA polymerase subunit sigma	*S. aurantiacus*	99%	284	WP_078621728.1
1239	mycothiol system anti-sigma-R factor	*S. aurantiacus*	98%	108	WP_016645393.1
1240	metal-dependent phosphohydrolase	*S. aureocirculatus*	89%	487	WP078965667.1
1241	metal-dependent phosphohydrolase	*S. silvensis*	91%	447	WP_058849656.1
1242	tetratricopeptide repeat domain-containing protein	*S. aurantiacus*	99%	326	WP_037662734.1
1243	peptide deformylase	*S. aurantiacus*	100%	216	WP_016644066.1
1244	putative Epi-isozizaene synthase terpene synthase	*S. aurantiacus* JA 4570 *Streptomyces* sp. ND90	99% 87%	322	EPH41005.1 BAP82226.1
1245	cytochrome P450	*S. aurantiacus*	99%	459	WP_016644064.1
1246	AraC family transcriptional regulator	*S. aurantiacus*	98%	329	WP_016644063.1
1247					
1248	putative Ribonucleoside-diphosphate reductase subunit M2 B	*S. aurantiacus*	99%	346	WP_016644062.1
1249	ribonucleoside-diphosphate reductase subunit alpha	*S. aurantiacus*	99%	814	WP_016644061.1
1250	symporter YodF	*S. aurantiacus*	99%	542	WP_016644060.1
1251	Protein of unknown function DUF3311	*Actinobacteria bacterium* OK006	75%	93	KPI21238.1
1252	DUF466 domain-containing protein	*S. aureocirculatus*	97%	66	WP_037687892.1
1253	carbon starvation protein A	*S. aurantiacus*	99%	722	WP_016644057.1
	Cluster 6 Phosphonate				
1366					
1367	HsdR family type I site-specific deoxyribonuclease	*Streptomyces* sp. HGB0020	72%	1054	WP_016437503.1
1368	site-specific integrase	*S. cellostaticus*	91%	271	WP_079057774.1
1369	DNA-binding protein	*Streptomyces* sp. CT34	91%	66	WP_043270748.1
1370	DNA repair protein RadA	*S. turgidiscabies*	92%	338	WP_059073140.1
1371					
1372					
1373					
1374	cytochrome P450	*S. flavovirens*	52%	414	WP_051748751.1
1375	cytochrome P450	*S. flavovirens*	45%	439	WP_051748750.1
1376	transcriptional regulator	*Streptomyces* sp. NRRL S-1813	33%	218	WP_051818737.1
1377	DNA-binding protein	*S. turgidiscabies*	54%	670	WP_006379070.1
1378	cytochrome P450	*S. flavovirens*	50%	324	WP_051748751.1
79					
80	ShlB/FhaC/HecB family hemolysin secretion/activation protein	*Pseudovibrio* sp. Ad26 Cov 60	47%	67	WP_063312315.1
81	transcriptional regulator	*S. griseus*	57%	68	WP_030720098.1
82	phosphoenolpyruvate phosphomutase	*S. griseus*	77%	296	WP_032769330.1
83	phosphonopyruvate decarboxylase	*S. monomycini*	58%	374	WP_030019523.1
84	DDE superfamily endonuclease	*Streptomyces* sp. Ag109_O5-10	69%	79	SED64923.1
85	AIPR protein	*Alloactinosynnema album*	64%	691	SDJ28210.1

86	restriction endonuclease subunit M	*S. griseus*	81%	713	WP_030753512.1
87	restriction endonuclease subunit S	*Streptomyces* sp. PBH53	76%	474	AKN72132.1
88	protein kinase	*Streptomyces* sp. NBRC 109706	79%	1452	WP_067225810.1
89					
90	DNA helicase HerA, contains HAS-barrel and ATPase domains	*Micromonospora pallida*	49%	1881	SCL38794.1
	Cluster 7 Siderophore				
1870					
1871					
1872	2,4-diaminobutyrate 4-aminotransferase	*S. aurantiacus*	96%	511	WP_037659335.1
1873	putative Rhizobactin siderophore biosynthesis protein RhbC	*S. aurantiacus* JA 4570	93%	697	EPH44073.1
1874	Lysine N-acyltransferase MbtK	*S. aurantiacus*	98%	258	WP_016641006.1
1875	IucA/IucC family siderophore biosynthesis protein	*S. aurantiacus*	98%	659	WP_016641005.1
1876	ATP-dependent helicase DinG like protein	*S. aurantiacus*	99%	667	WP_016641004.1
1877	epressor LexA	*Streptomyces* sp. NRRL B-1347	98%	261	WP_030680474.1
1878	transcriptional regulator NrdR	*S. aurantiacus*	99%	180	WP_037659333.1
1879	vitamin B12-dependent ribonucleotide reductase	*S. aurantiacus*	99%	963	WP_016641001.1
	Cluster 8 NRPS				
1878	transcriptional regulator NrdR	*S. aurantiacus*	99%	180	WP_037659333.1
1879	vitamin B12-dependent ribonucleotide reductase	*S. aurantiacus*	99%	963	WP_016641001.1
1880	cAMP-binding protein 2	*S. aurantiacus*	97%	175	WP_016640999.1
1881	arylamine N-acetyltransferase	*S. aurantiacus*	95%	273	WP_016641000.1
1882	DUF4937 domain-containing protein	*S. aurantiacus*	99%	201	WP_037659328.1
1883	TetR/AcrR family transcriptional regulator	*S. aurantiacus*	95%	237	WP_016640997.1
1884	ABC transporter ATP-binding protein	*S. aurantiacus*	98%	536	WP_037659327.1
1885	histidine phosphatase family protein	*S. aurantiacus*	99%	219	WP_016640994.1
1886	ADP-ribosylglycohydrolase family protein	*S. aurantiacus*	98%	302	WP_016640993.1
1887	putative Puromycin resistance protein pur8	*S. aurantiacus* JA 4570	99%	496	EPH44058.1
1888	TetR family transcriptional regulator	*S. aurantiacus*	99%	204	WP_037659323.1
1889	ribonuclease HII	*S. aurantiacus*	99%	233	WP_016640990.1
1890					
1891					
1892	ATP-dependent DNA helicase RecQ	*S. aurantiacus*	99%	719	WP_016640986.1
1893	2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase	*S. aurantiacus*	98%	267	WP_016640985.1
1894	aminodeoxychorismate synthase	*S. uncialis*	90%	398	WP_073788175.1
1895	putative 2,3-dihydroxybenzoate-AMP ligase	*S. aurantiacus* JA 4570	98%	570	EPH43227.1
1896	2,3-dihydro-2,3-dihydroxybenzoate synthetase	*S. aurantiacus*	91%	222	WP_037660277.1
1897	acyl carrier protein	*Streptomyces*. sp. TSRI0281	87%	82	WP_073719042.1
1898	siderophore-interacting protein	*S. aurantiacus*	98%	282	WP_016641880.1
1899	iron(3+)-hydroxamate-binding protein YxeB ABC transporter substrate-binding protein	*S. aurantiacus* *Streptomyces* sp. NRRL B-1347	99% 86%	331	WP_016641879.1 WP_030682954.1
1900	iron ABC transporter permease ABC transporter permease	*S. aurantiacus* *S. silaceus*	99% 86%	340	WP_016641878.1 WP_076687382.1
1901	ferric anguibactin ABC transporter permease	*S. aurantiacus*	99%	332	WP_016641877.1
1902	ABC transporter ATP-binding protein	*S. aurantiacus*	99%	261	WP_016641876.1

1903	non-ribosomal peptide synthetase	*S. luteocolor*	85%	3466	WP_069883640.1
1904	MFS transporter	*S. aurantiacus*	98%	476	WP_016641872.1
1905	3-deoxy-7-phosphoheptulonate synthase	*S. aurantiacus*	95%	385	WP_078621315.1
1906	DUF2470 domain-containing protein	*Streptomyces*. sp. CB02488	62%	255	WP_073735179.1
1907	signal transduction histidine kinase	*Novosphingobium* sp. MD-1	40%	60	GAO53879.1
1908	DUF4192 domain-containing protein	*S. aurantiacus*	98%	488	WP_016641867.1
1909	glycogen debranching protein	*S. aurantiacus*	99%	701	WP_078621313.1
1910	NUDIX domain-containing protein	*S. aurantiacus*	100%	253	WP_016641866.1
1911	ABC transporter	*S. kanamyceticus*	77%	637	WP_055544933.1
1912	FadR family transcriptional regulator	*S. aurantiacus*	99%	289	WP_016641862.1
1913	RNA polymerase sigma factor	*S. aurantiacus*	99%	514	WP_016641861.1
1914	serine protease	*S. aurantiacus*	99%	274	WP_037660276.1
1915					
1916	DNA topoisomerase IV subunit B	*S. aurantiacus*	99%	707	WP_016641856.1
1917	DUF485 domain-containing protein	*S. aurantiacus*	94%	168	WP_016641854.1
1918	cation acetate symporter	*S. aurantiacus*	99%	530	WP_037660273.1
	Cluster 9 T1 PKS				
1957	potassium-transporting ATPase subunit A	*S. aurantiacus*	100%	554	WP_037657573.1
1958	potassium-transporting ATPase B chain	*S. aurantiacus*	99%	704	WP_016639108.1
1959	potassium-transporting ATPase C chain	*S. aurantiacus*	96%	239	WP_016639109.1
1960	sensor histidine kinase	*S. kanamyceticus*	91%	848	WP_055551025.1
1961	DNA-binding response regulator	*S. aurantiacus*	100%	227	WP_016645789.1
1962	DNA-binding protein	*Streptomyces*. sp. NRRL B-1347	98%	116	WP_078869048.1
1963	DUF3159 domain-containing protein	*Streptomyces*. sp. CNR698	75%	243	WP_027734030.1
1964	trk system potassium uptake protein trkA	*S. aurantiacus*	100%	220	WP_016645786.1
1965	trk system potassium uptake protein trkA	*S. aurantiacus*	99%	217	WP_016645785.1
1966					
1967	transcriptional regulator	*S. aurantiacus*	100%	121	WP_037664712.1
1968	putative membrane protein	*S. aurantiacus* JA 4570	98%	757	EPH39263.1
1969	amino acid permease YdaO DNA-binding protein	*S. aurantiacus* *Streptomyces*. sp. NRRL B-1347	99% 95%	709	WP_016645781.1 WP_030682413.1
1970	class I SAM-dependent RNA methyltransferase	*S. aurantiacus*	99%	440	WP_016645780.1
1971	PucR family transcriptional regulator	*Streptomyces*. sp. 769	63%	88	WP_039638427.1
1972	putative Lactate 2-monooxygenase	*S. aurantiacus* JA 4570	86%	736	EPH39258.1
1873	putative Phenolphthiocerol synthesis polyketide synthase type I Pks15/1	*S. aurantiacus*	96%	2178	WP_016645776.1
1974	polyketide synthase Acyl transferase domain-containing protein	*Saccharopolyspora erythraea* *Streptomyces* sp. 2112.2	47% 46%	2862	WP_011873765.1 SEE91982.1
1975	cytochrome P450 *S. aurantiacus*	*S. aurantiacus*	97%	387	WP_037664702.1
1976	RNA polymerase sigma-C factor	*S. aurantiacus*	96%	198	WP_016645773.1
1977					
1978	arabinose ABC transporter permease	*S.* sp. Root369	81%	433	KQV94405.1
1979	sugar ABC transporter permease	*S.* sp. XY332	63%	459	WP_053788035.1
1980	acyl-CoA dehydrogenase *S. aurantiacus*	*S. aurantiacus*	96%	402	WP_078620988.1
1981	acyl-CoA transferase	*S. silvensis*	75%	657	WP_058849428.1
1982	Triostin synthetase I	*S. aurantiacus*	97%	559	WP_016639365.1
1983	ABC transporter substrate-binding protein	*S. silvensis*	84%	274	WP_058849489.1
1984	metal ABC transporter permease	*S. aurantiacus*	97%	293	WP_016639363.1
1985	putative Lipoprotein-releasing system ATP-binding protein LolD	*S. aurantiacus* JA 4570	98%	212	EPH45726.1
1986	putative lipoprotein	*S. aurantiacus* JA 4570	97%	174	EPH45725.1
1987	putative Cation-transporting P-type ATPase B	*S. aurantiacus* JA 4570	92%	788	EPH45724.1
1988					

1989	cell division protein SepF	*S. aurantiacus*	99%	127	WP_016639357.1
1990	biotin transporter BioY	*S. aureocirculatus*	93%	394	WP_051853014.1
1991	nucleotide pyrophosphohydrolase	*S. aurantiacus*	97%	107	WP_016639355.1
	Cluster 10 Bacteriocin				
2064	protoporphyrinogen oxidase	*S. kanamyceticus*	79%	490	WP_055544609.1
2065	chlorite dismutase *S.* indicus		96%	239	SDK28577.1
2066	RNHCP domain-containing protein	*S. aurantiacus*	99%	124	WP_016643865.1
2067	alpha/beta hydrolase	*S. aurantiacus*	97%	537	WP_016643866.1
2068	TIGR04222 domain-containing membrane protein	*S. aurantiacus*	98%	261	WP_016643867.1
2069	Tat pathway signal sequence domain protein	*S. aurantiacus*	99%	261	WP_037662494.1
2070	endonuclease	*S. aureocirculatus*	84%	464	WP_051853385.1
2071					
2072					
2073	peptidyl-tRNA hydrolase	*S. olivochromogenes*	84%	258	WP_067375498.1
2074	dienelactone hydrolase family protein	*S. aurantiacus*	97%	244	WP_037662497.1
2075					
2076	polysaccharide deacetylase	*S. silaceus*	76%	310	WP_076684012.1
	Cluster 11 Lassopeptide				
2221	putative HTH-type transcriptional regulator YobV	*S. aurantiacus* JA 4570	96%	335	EPH46501.1
2222	MFS transporter	*S. aurantiacus*	96%	525	WP_016638606.1
2223	gfo/Idh/MocA family oxidoreductase	*S. aurantiacus*	95%	244	WP_016638605.1
2224					
2225					
2226					
2227					
2228	putative restriction endonuclease	*Gordonia polyisoprenivorans* NBRC 16320 = JCM 10675	46%	620	GAB22404.1
2229					
2230					
2231	asparagine synthase	*S. aurantiacus*	97%	607	WP_037657108.1
2232	PqqD family protein	*S. aurantiacus*	99%	84	WP_016638601.1
2233	polyketide beta-ketoacyl synthase	*S.* sp. NBRC 110027	81%	144	WP_042159724.1
2234	putative ABC transporter ATP-binding protein	*S. aurantiacus* JA 4570	97%	618	EPH46493.1
2235	putative Magnesium transport protein CorA	*S. aurantiacus* JA 4570	99%	337	EPH46461.1
2236	winged helix DNA-binding domain-containing protein	*S. aurantiacus*	92%	409	WP_037657093.1
2237	LysE family translocator	*S. aurantiacus*	98%	217	WP_016638565.1
2238	helix-turn-helix transcriptional regulator	*S. aurantiacus*	93%	304	WP_016638564.1
2239	acyl-CoA synthetase AMP-dependent synthetase	*S. aurantiacus* *S.* sp. NRRL B-1347	98% 92%	569	WP_016638563.1 WP_030667345.1
2240	acyl-CoA synthetase YngI AMP-binding protein	*S. aurantiacus* *S. aureocirculatus*	97% 92%	556	WP_016638562.1 WP_030564861.1
2241					
	Cluster 12 Terpene				
2710	NACHT domain-containing protein ATP-binding protein	*S. aurantiacus* *S. aureocirculatus*	99% 87%	1018	WP_016644096.1 WP_030560821.1
2711	metallophosphoesterase	*S. aurantiacus*	99%	286	WP_052029339.1
2712	putative 2,5-dichloro-2,5-cyclohexadiene-1,4-diol dehydrogenase NAD(P)-dependent dehydrogenase, short-chain alcohol dehydrogenase family	*S. aurantiacus* JA 4570 *Streptomyces*. sp. cf386	98% 73%	284	EPH40982.1 SDO78283.1
2713	galactonate dehydratase	*S. aurantiacus*	99%	381	WP_016644100.1

2714	cytochrome bc complex cytochrome b subunit	*S. aurantiacus*	99%	547	WP_016644101.1
2715	tetratricopeptide repeat protein	*S. aurantiacus*	97%	818	WP_037662753.1
2716					
2717	putative Germacradienol/geosmin synthase	*S. aurantiacus* JA 4570	99%	738	EPH40987.1
2718	putative HTH-type transcriptional regulator CueR	*S. aurantiacus*	79%	193	WP_016644105.1
2719	NADP-dependent oxidoreductase	*S. aurantiacus*	97%	349	WP_037662755.1
2720	Predicted arabinose efflux permease, MFS family	*Saccharopolyspora antimicrobica*	70%	392	SFN36552.1
2721	TetR family transcriptional regulator	*S. silvensis*	84%	298	WP_058851389.1
2722	3-keto-5-aminohexanoate cleavage protein NADPH:quinone reductase	*S. aurantiacus* *S. silaceus*	99% 96%	295	WP_037663535.1 WP_076683272.1
2723	L-carnitine dehydrogenase 3-hydroxybutyryl-CoA dehydrogenase	*S. aurantiacus* *S. silaceus*	95% 94%	325	WP_016644767.1 WP_055702492.1
2724	4-hydroxybenzoyl-CoA thioesterase L-carnitine dehydrogenase	*S. aureocirculatus* *S. aurantiacus*	90% 95%	166	WP_030560812.1 WP_016644768.1
2725	6-phospho-beta-glucosidase	*S. aurantiacus*	99%	446	WP_037663537.1
2726					
2727	FBP domain-containing protein	*S. aurantiacus*	100%	167	WP_016645233.1
2728	NADPH:quinone reductase	*S. aurantiacus*	97%	306	WP_037664082.1
	Cluster 13 Terpene				
2811	putative Cation efflux system protein CzcD	*S. aurantiacus* JA 4570	100%	310	EPH45790.1
2812	UDP-glucose 4-epimerase GalE	*S. aurantiacus*	99%	326	WP_016639281.1
2813	transferase	*S. aurantiacus*	95%	623	WP_037657720.1
2814	mannose-1-phosphate guanylyltransferase 1 nucleotide sugar-1-phosphate transferase Choline kinase	*S. aurantiacus* *Streptomyces* sp. NRRL S-920 *S. indicus*	99% 98% 91%	245	WP_016639279.1 WP_030781248.1 SDJ49279.1
2815	glycerol-1-phosphate dehydrogenase dehydrogenase	*S. aurantiacus* *Streptomyces* sp. NRRL B-1347	99% 97%	354	WP_016639278.1 WP_030684554.1
2816	CDP-alcohol phosphatidyltransferase	*Streptomyces* sp. KS_5	88%	227	SEB74082.1
2817	glycosyl transferase	*S. aurantiacus*	99%	300	WP_037657733.1
2818	putative O-antigen export system permease protein RfbA ABC transporter	*S. aurantiacus* JA 4570 *S. aureocirculatus*	99% 96%	298	EPH45783.1 WP_030560732.1
2819	putative Teichoic acids export ATP-binding protein TagH	*S. aurantiacus* JA 4570	99%	270	EPH45782.1
2820	squalene synthase HpnC	*S. aurantiacus*	99%	300	WP_037657716.1
2821	squalene synthase HpnD	*S. aurantiacus*	99%	316	WP_037657714.1
2822	zeta-carotene desaturase phytoene dehydrogenase	*S. aurantiacus* *Streptomyces* sp. NRRL B-1347	99% 90%	468	WP_016639271.1 WP_030688854.1
2823	putative (2E,6E)-farnesyl diphosphate synthase dimethylallyltranstransferase	*S. aurantiacus* *S. silvensis*	99% 97%	374	WP_016639270.1 WP_058851462.1
2824	squalene-hopene cyclase	*S. aurantiacus*	99%	664	WP_016639269.1
2825	1-hydroxy-2-methyl-2-butenyl 4-diphosphate reductase	*S. silaceus*	88%	216	WP_055697309.1
2826	hopanoid biosynthesis associated radical SAM protein HpnH	*S. aurantiacus*	100%	339	WP_016639267.1
2827	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	*S. kanamyceticus*	96%	388	WP_055545784.1
2828	putative 1-deoxy-D-xylulose-5-phosphate synthase 1	*S. aurantiacus* JA 4570	96%	686	EPH42220.1
2829	aspartate aminotransferase family protein	*S. aurantiacus*	99%	464	WP_016642847.1
2830	XRE family transcriptional regulator	*S. aurantiacus*	99%	210	WP_016642846.1
2831	small hydrophobic protein	*Actinobacteria bacterium* OK006	63%	62	KPI09663.1
2832	protein-tyrosine-phosphatase	*S. aurantiacus*	99%	265	WP_016642844.1

2833	alpha-galactosidase C	*S. aurantiacus*	96%	700	WP_016642842.1
	Cluster 14 T1 PKS				
2854	sugar-binding protein	*S. aureocirculatus*	89%	1152	WP_030572827.1
2855	alpha/beta hydrolase *S. aurantiacus* lysophospholipase	*S. aurantiacus* *S. aureocirculatus*	96% 82%	343	WP_016641398.1 WP_030572829.1
2856					
2857	MFS transporter	*S. aurantiacus*	96%	516	WP_016641400.1
2858	transketolase	*Streptomyces* sp. NRRL S-646	82%	304	WP_030936098.1
2859	Ferredoxin fas2 transketolase	*S. aurantiacus* *S. silvensis*	99% 93%	233	WP_016641401.1 WP_058847312.1
2860	Bacilysin biosynthesis protein BacA	*S. aurantiacus*	99%	210	WP_016641402.1
2861					
2862	nuclear transport factor 2 family protein ketosteroid isomerase	*S. aurantiacus* uncultured bacterium 16	100% 45%	144	WP_016641404.1 ALV86410.1
2863	branched-chain amino acid aminotransferase	*S. aurantiacus*	98%	313	WP_016641405.1
2864	4-azaleucine resistance probable transporter AzlC branched-chain amino acid permease	*Nocardioides exalbidus* *Marmoricola* sp. Leaf446	40% 40%	246	SED14003.1 WP_056544661.1
2865	AzlD domain-containing protein	*S. aurantiacus*	99%	105	WP_016641407.1
2866	AfsR/SARP family transcriptional regulator	*S. silvensis*	83%	1096	WP_058847306.1
2867					
2868	nuclear transport factor 2 family protein ketosteroid isomerase	*Streptomyces* sp. ERV7 *Actinophytocola xanthii*	75% 42%	141	WP_067161415.1 WP_075124451.1
2869	histidine kinase	*S. iranensis*	58%	184	WP_044568772.1
2870	transposase	*S. silvensis*	91%	411	WP_058847515.1
2871	antibiotic biosynthesis protein	*S. silvensis*	94%	231	WP_079086215.1
72	protease synthase and sporulation protein PAI 2	*S. aurantiacus*	99%	214	WP_016641415.1
2873	L-ornithine 5-monooxygenase L-lysine 6-monooxygenase	*S. aurantiacus* *S. silvensis*	99% 89%	436	WP_016641416.1 WP_058847301.1
2874	MbtH family protein	*S. aurantiacus*	100%	74	WP_016641417.1
2875	bifunctional ornithine acetyltransferase/N-acetylglutamate synthase glutamate N-acetyltransferase	*S. aurantiacus* *S. wuyuanensis*	98% 85%	384	WP_016641418.1 SDM84429.1
2876					
2877	aspartate aminotransferase family protein acetylornithine aminotransferase	*S. aurantiacus* *S. aureocirculatus*	99% 93%	411	WP_016641420.1 WP_037684869.1
2878	Double check			233	
2879	putative membrane protein YdfJ	*S. aurantiacus*	98%	752	WP_016641423.1
2880	D-glutamate deacylase N-acyl-D-aspartate/D-glutamate deacylase	*S. aurantiacus* *Amycolatopsis saalfeldensis*	97% 60%	525	WP_078621258.1 SEO47913.1
2881	exopolyphosphatase / guanosine-5'-triphosphate,3'-diphosphate pyrophosphatase	*Streptomyces* sp. cf124	70%	336	SFN31148.1
2882	McpG MarG	*S. longispororuber* Cov 84 *Streptomyces* sp. CNQ-617	99% 53%	413	AEL16995.1 AHF22859.1
2883	McpH MarH	*S. longispororuber* *Streptomyces* sp. CNQ-617	99% 70%	944	AEL16993.1 AHF22858.1
2884	putative Demethylspheroidene O-methyltransferase SAM-dependent methyltransferase MarI McpI	*S. aurantiacus* *S. silvensis* *Streptomyces* sp. CNQ-617 *S. longispororuber* Cov 33	99% 88% 66% 99%	381	WP_016641428.1 WP_058847294.1 AHF22857.1 AEL16994.1

2885	thioesterase MarJ	*S. aurantiacus* *Streptomyces* sp. CNQ-617	96% 63%	281	WP_016641430.1 AHF22856.1
2886	MarL	*S.* sp. CNQ-617 Cov 85	61%	2671	AHF22854.1
2887	D-alanine--poly(phosphoribitol) ligase MarM	*S. aurantiacus* *Streptomyces* sp. CNQ-617	96% 63%	560	WP_052028931.1 AHF22853.1
2888	8-amino-7-oxononanoate synthase-amino-3-ketobutyrate CoA ligase MarN	*S. aurantiacus* *Streptomyces* sp. CNQ-617	98% 71%	616	WP_016638827.1 AHF22852.1
2889	D-alanine--poly(phosphoribitol) ligase subunit 2 MarO	*S. aurantiacus* *Streptomyces* sp. CNQ-617	99% 57%	97	WP_016638826.1 AHF22851.1
2890	acyl carrier protein MarQ	*S. aurantiacus* *Streptomyces* sp. CNQ-617	100% 59%	93	WP_016638825.1 AHF22849.1
2891	ketoacyl-ACP synthase III MarP	*S. aurantiacus* *Streptomyces* sp. CNQ-617	99% 70%	343	WP_037657312.1 AHF22850.1
2892	putative 3-oxoacyl-acyl-carrier-protein synthase 2 MarR	*S. aurantiacus* *Streptomyces* sp. CNQ-617	99% 77%	436	WP_016638822.1 AHF22848.1
2893	MarT	*Streptomyces* sp. CNQ-617	64%	296	AHF22847.1
2894	MarU	*Streptomyces* sp. CNQ-617	60%	253	AHF22846.1
2895	MarV RedV protein	*Streptomyces* sp. CNQ-617 *S. coelicolor*	48% 52%	443	AHF22845.1 WP_011030511.1
2896	MarZ	*Streptomyces* sp. CNQ-617	50%	224	AHF22844.1
2897	RedY protein MarY	*S. aureocirculatus* *Streptomyces* sp. CNQ-617	75% 61%	121	WP_051853027.1 AHF22843.1
2898	acyl-CoA dehydrogenase RedW *S.* sp. CNQ-617	*S. aureocirculatus* *S. aurantiacus*	89% 75%	406	WP_030568386.1 AHF22842.1
2899	putative Phthiocerol/phenolphthiocerol synthesis polyketide synthase type I PpsB MarX	*S. aurantiacus* JA 4570 *Streptomyces* sp. CNQ-617	96% 54%	1183	EPH46241.1 AHF22841.1
2900	putative transcriptional regulator RedD MarD	*S. aurantiacus* JA 4570 *Streptomyces* sp. CNQ-617	98% 69%	261	EPH39961.1 AHF22840.1
2901	TetR family transcriptional regulator	*S. aurantiacus*	97%	213	WP_037663936.1
2902	ABC transporter ATP-binding protein	*S. aurantiacus*	97%	215	WP_016645110.1
2903	integral membrane protein	*S. griseoaurantiacus* M045	73%	265	EGG45266.1
2904	cysteine desulfurase	*S. aurantiacus*	96%	460	WP_016645108.1
	Cluster 15 T1 PKS				
2943	MFS transporter	*S. aurantiacus*	93%	441	WP_016645434.1
2944	NADPH:quinone reductase	*S. silvensis*	90%	380	WP_058847256.1
2945	TetR family transcriptional regulator	*S. aurantiacus*	98	238	WP_016645432.1
2946	N-acetyltransferase	*S. silvensis*	82	327	WP_058847254.1
2947	glycosyl hydrolase	*S. silvensis*	89	772	WP_079086204.1
2948	ABC transporter ATP-binding protein	*S. silvensis*	94	638	WP_058847501.1
2949	TetR family transcriptional regulator	*Streptomyces* sp. NRRL B-1347	84%	191	WP_030681573.1
2950	4'-phosphopantetheinyl transferase	*Streptomyces* sp. CNS654	74%	270	WP_030357188.1
2951	dolichol-phosphate mannosyltransferase	*Streptomyces* sp. NRRL WC-3618	92%	411	KOV63872.1
2952	helix-turn-helix transcriptional regulator	*Streptomyces* sp. NRRL WC-3618	91%	943	WP_053744793.1

2953	ABC transporter ATP-binding protein	*S. albus* Cov 52	47%	252	WP_037681692.1
2954	amine oxidase NAD(P)/FAD-dependent oxidoreductase	*Streptomyces* sp. NRRL WC-3618 *Frankia discariae*	95% 69%	542	WP_053744791.1 WP_018504767.1
2955	ABC transporter	*Streptomyces* sp. NRRL *Streptomyces* C-3618	97%	247	WP_053744790.1
2956	daunorubicin ABC transporter ATPase	*Streptomyces* sp. NRRL WC-3618	95%	314	KOV63897.1
2957	modular polyketide synthase Acyl transferase domain-containing protein	*Streptomyces* sp. RK95-74 *S. melanosporofaciens*	57% 57%	5219	BAW35608.1 SEB92911.1
2958	Acyl transferase domain-containing protein	*S. melanosporofaciens*	59%	3277	SEB92911.1
2959	Acyl transferase domain-containing protein eta-ketoacyl synthase	*S. melanosporofaciens* *S.* iranensis	55% 54%		SEB92911.1 CDR09745.1
2960	beta-ketoacyl synthase Acyl transferase domain-containing protein	*Streptomyces* sp. NBRC 110611 *Streptomyces* sp. yr375	54% 54%	5056	WP_066934984.1 SES19490.1
2961	polyketide synthase beta-ketoacyl synthase	*Streptomyces* sp. SPMA113 *Streptomyces* sp. NBRC 109436	56% 55%	5264	WP_079152784.1 WP_079150001.1
2962	Phosphopantetheine attachment site 3-ketoacyl-ACP reductase	*S. melanosporofaciens* *S. hygroscopicus*	56% 55%	3428	SED16705.1 WP_078641494.1
2963	mycolactone core lactone polyketide synthase MlsA1 soraphen polyketide synthase B polyene macrolide polyketide synthase	*Streptomyces* sp. DI166 *Sorangium cellulosum* *Micromonospora humi*	54% 50% 54%	5345	SBT94599.1 AAA79984.2 SCG63874.1
2964	type I polyketide synthase 3-ketoacyl-ACP reductase	*S. ruber* *Streptomyces* sp. PRh5	83% 57%	4233	WP_051834930.1 WP_078559686.1
2965	4-amino-4-deoxy-L-arabinose transferase	*Amycolatopsis marina*	51%	536	SFB55897.1
2966	4-coumarate--CoA ligase *S. aurantiacus*	*S. aurantiacus*	100%	478	WP_016642108.1
2967	malonyl CoA-ACP transacylase	*S. aurantiacus*	96%	337	WP_016642107.1
2968	cytochrome P450	*S. aurantiacus*	99%	423	WP_037660518.1
2969	proclavaminate amidinohydrolase agmatinase	*S. aurantiacus* *S.* sp. NRRL B-1347	99% 98%	372	WP_016642105.1 WP_030688343.1
2970	thioesterase	*S. aurantiacus*	100%	249	WP_016642104.1
2971					
2972	two-component sensor histidine kinase	*S. aurantiacus*	99%	442	WP_052029206.1
2073	DNA-binding response regulator	*S. aurantiacus*	100%	220	WP_037663178.1
2974	two-component sensor histidine kinase	*S. aurantiacus*	99%	424	WP_052029206.1
2975	membrane protein	*Streptomyces* sp. NRRL B-1347	75%	184	WP_030684489.1
2976	membrane protein	*Streptomyces* sp. RSD-27	43%	512	WP_042816496.1
2977	putative membrane protein	*S. aurantiacus*	99%	729	WP_016642726.1
2978					
2979					
2980	putative L-amino-acid oxidase	*S. aurantiacus* JA 4570	96%	514	EPH42339.1
2981	cold-shock protein DNA-binding protein	*S. aurantiacus* *Streptomyces* sp. NRRL B-1347	97% 78%	141	WP_016642721.1 WP_030684496.1
	Cluster 16 Lassopeptide				
3259	nickel import ATP-binding protein NikE ABC transporter	*S. aurantiacus* *Streptomyces* sp. NRRL B-1347	93% 93%	219	WP_016638977.1 WP_030675447.1
3260	LysR family transcriptional regulator DNA-binding transcriptional regulator, LysR family	*S. aurantiacus* *S.* melanosporofaciens	97% 83%	269	WP_016638978.1 SED35903.1
3261	NmrA family transcriptional regulator putative Quinone oxidoreductase 2	*S. aurantiacus* *S. aurantiacus* JA 4570	96% 96%	299	WP_037657459.1 EPH46111.1
3262	DeoR/GlpR transcriptional regulator	*S. aurantiacus*	96%	262	WP_016638980.1

3263	MFS transporter	*S. aurantiacus*	97%	444	WP_016638981.1
3264	carbohydrate kinase	*S. aurantiacus*	98%	514	WP_016638982.1
3265	xylulose kinase	*S. aurantiacus*	95%	507	WP_016638983.1
3266	histidine phosphatase family protein	*S. aurantiacus*	99%	204	WP_016638984.1
3267	D-3-phosphoglycerate dehydrogenase	*S. aurantiacus*	99%	346	WP_016638985.1
3268	Hypo				
3269	asparagine synthase	*S. aurantiacus*	96%	630	WP_052028944.1
3270	PqqD family protein	*S. aurantiacus*	98%	85	WP_037657425.1
3271	polyketide beta-ketoacyl synthase	*S.* rubellomurinus	63%	139	WP_045710139.1
3272	Hypo				
3273	glyoxalase/bleomycin resistance/dioxygenase family protein	*Mycobacterium colombiense*	92%	122	WP_064879765.1
3274	MarR family transcriptional regulator	*S. aureocirculatus*	95%	150	WP_030570658.1
3275	Glycerophosphoryl diester phosphodiesterase family protein	*Streptomyces* sp. DI166	73%	302	SBT94407.1
3276	Hypo				
3277	Hypo				
3278	Hypo				
3279	UDP-glucose 4-epimerase	*S. aurantiacus*	100%	118	WP_037657429.1
3280	Hypo				
3281	Hypo				
3282	peptidase M4	*S. atriruber*	67%	232	WP_079074477.1
3283	putative UPF0678 fatty acid-binding protein-like protein	*S. aurantiacus* JA 4570	96%	179	EPH46136.1
	Cluster 17 Bacteriocin				
3314	NHLP family bacteriocin export ABC transporter permease/ATPase subunit	*S. aurantiacus*	97%	958	WP_016640755.1
3315	NHLP family bacteriocin export ABC transporter peptidase/permease/ATPase subunit ABC transporter	*S. aurantiacus* *Streptomyces* sp. WM6378	99% 90%	761	WP_078621181.1 KOU38317.1
3316	HlyD family secretion protein	*Streptomyces* sp. NBRC 110027	72%	268	WP_042152488.1
3317	type A2 lantipeptide	*Streptomyces* sp. 13-12-16	66%	75	WP_085572576.1
3318	membrane protein phosphoesterase PA-phosphatase related protein	*S. lydicus* *S. iranensis*	82% 84%	240	WP_046929921.1 CDR13627.1
3319	glutamate-1-semialdehyde 2,1-aminomutase	*S. aurantiacus*	97%	441	WP_016640760.1
3320	dTDP-4-dehydrorhamnose 3,5-epimerase	*Streptomyces* sp. RTd22	93%	183	WP_063726508.1
3321	membrane protein DTDP-4-DEHYDRORHAMNOSE 3,5-EPIMERASE	*S. aurantiacus* *S. venezuelae*	96% 86%	398	WP_037659037.1 CUM36249.1
3322	glycosyl transferase	*S. regalis*	92%	318	WP_062703767.1
	Cluster 18 Lassopeptide				
3441	ABC transporter ATP-binding protein	*S. aurantiacus*	97%	627	WP_016640856.1
3442	aminodeoxychorismate lyase	*S. aureocirculatus*	88%	238	WP_030572756.1
3443	serine protease	*S. aurantiacus*	98%	422	WP_037659087.1
3444	peptidase M4	*S. aurantiacus*	97%	924	WP_016640861.1
3445	sigma-70 family RNA polymerase sigma factor	*S. aureocirculatus*	90%	301	WP_063760850.1
3446	Hypo				
3447	Ls3A-pre-pro-peptide (plasmid)	Cov 61 *S. leeuwenhoekii*	66%	47	CQR59431.1
3448	asparagine synthase asparagine synthase (glutamine-hydrolysing)	*S. aurantiacus* *Streptomyces* sp. SceaMP-e96	94% 54%	669	WP_016640865.1 SCK29539.1
3449	PqqD family protein	*S. aurantiacus*	97%	87	WP_016640866.1
3450	Transglutaminase-like superfamily protein	*Streptomyces* sp. SceaMP-e96	76%	145	SCK29524.1
3451	NAD(P)/FAD-dependent oxidoreductase	*S. aurantiacus*	98%	366	WP_016640868.1

3452	alkanesulfonate monooxygenase	*S. aurantiacus*	99%	402	WP_016640869.1
3453	putative aliphatic sulfonates-binding protein ABC transporter substrate-binding protein	*S. aurantiacus* *S. aureocirculatus*	98% 94%	353	WP_016640870.1 WP_030572742.1
3454	aliphatic sulfonate ABC transporter ATP-binding protein 1	*S. aurantiacus*	98%	280	WP_016640871.1
3455	ABC transporter permease	*S. aurantiacus*	97%	297	WP_016640872.1
3456	putative UPF0047 protein	*S. aurantiacus* JA 4570	99%	140	EPH44304.1
3457	serine/threonine protein phosphatase	*S. aurantiacus*	99%	248	WP_037659169.1
3458	haloacid dehalogenase	*S. aurantiacus*	94%	303	WP_016640876.1
3459	phosphoribosyltransferase	*S. aurantiacus*	96%	467	WP_016640875.1
3460	pyridine nucleotide-disulfide oxidoreductase putative NADPH-ferredoxin reductase FprA	*S. aurantiacus* *S. aurantiacus* JA 4570	99% 99%	454	WP_037659091.1 EPH44308.1
	Cluster 19 Other				
3589	putative ATP-dependent Clp protease proteolytic subunit 2	*S. aurantiacus* JA 4570	100%	229	EPH45115.1
3590	deaminase/reductase	*S. fulvoviolaceus*	95%	192	WP_030620805.1
3591	Hyp				
3592	putative amino acid permease YdaO amino acid/polyamine/organocation transporter, APC superfamily	*S. aurantiacus* *Streptomyces* sp. DvalAA-43	98% 78%	682	WP_016639878.1 SCE13540.1
3593	AAC(3) family N-acetyltransferase aminoglycoside N(3)-acetyltransferase	*S. aurantiacus* Herbidospora mongoliensis	94% 63%	281	WP_016639879.1 WP_066367492.1
3594	DDE transposase IS5/IS1182 family transposase	*S. lydicus* *S. venezuelae*	88% 85%	276	WP_046929989.1 WP_079037206.1
3595	Hypo				
3596	DUF3224 domain-containing protein	*S. carpinensis*	75%	103	WP_086730171.1
3697	transcriptional regulator DNA-binding protein	*Streptomyces* sp. MMG1533 *S. pristinaespiralis*	68% 63%	185	WP_053749823.1 ALC21243.1
3598	Hypo				
3599	protein translocase subunit SecDF SecD/SecF fusion protein	*S. aurantiacus* *Streptomyces* sp. DI166	98% 76%	789	WP_016639882.1 SBT94514.1
3600	Hypo				
3601	IgA Peptidase M64	*Streptomyces* sp. SceaMP-e96	65%	440	SCK08638.1
3602	transcriptional regulator	*S. aurantiacus*	99%	536	WP_016639885.1
3603	Hypo				
3604	SAM-dependent methyltransferase	*S. kanamyceticus*	77%	61	WP_055549705.1
3605	putative HTH-type transcriptional repressor DasR	*S. aurantiacus* JA 4570	100%	286	EPH45128.1
3606	phosphogluconate dehydrogenase	*S. aurantiacus*	98%	256	WP_037658248.1
3607	putative branched-chain-amino-acid aminotransferase 3-hydroxyisobutyrate dehydrogenase	*S. aurantiacus* JA 4570 *Nonomuraea solani*	98% 58%	315	EPH45131.1 SEH03286.1
3608	putative transporter YbhF	*S. aurantiacus* JA 4570	98%	307	EPH45132.1
3609	alanine--anticapsin ligase ATP-grasp domain-containing protein	*S. aurantiacus* *Streptomyces* sp. WM6368	98% 51%	443	WP_016639891.1 WP_053702845.1
3610	dihydroneopterin 2',3'-cyclic phosphate phosphodiesterase	*S. aurantiacus*	99%	201	WP_016639892.1
3611	Hypo				
3612	glutamate-5-semialdehyde dehydrogenase	*S. aurantiacus*	98%	414	WP_016639894.1
3613	aspartate aminotransferase family protein	*S. aurantiacus*	98%	406	WP_016639895.1
3614	peptidase	*Streptomyces* sp. WM6368	68%	258	WP_053702842.1
3615	ATP/GTP-binding protein	*Streptomyces* sp. CB01249	48%	97	WP_073864110.1
3616					

3617	hydrolase	*Streptomyces* sp. WAC04657	74%	59	WP_062757634.1
3618	carboxymuconolactone decarboxylase family protein	*S. aurantiacus*	98%	159	WP_037658252.1
3619	DUF4440 domain-containing protein	*Microtetraspora malaysiensis*	87%	133	WP_084259219.1
3620	NmrA/HSCARG family protein	*S. aurantiacus*	95%	280	WP_016639901.1
3621	TetR family transcriptional regulator	*S. aurantiacus*	95%	224	WP_037658307.1
3622	Hypo				
3623	Hypo				
3624	Hypo				
3625	Hypo				
3626	Hypo				
3627	putative Cadmium, cobalt and zinc/H(+)-K(+) antiporter	*S. aurantiacus* JA 4570	96%	224	EPH45148.1
3628	histidine phosphatase family protein	*S. aurantiacus*	98%	218	WP_016639907.1
3629	lysine transporter LysE	*S. aurantiacus*	97%	218	WP_078621086.1
3630	putative ABC transporter extracellular-binding protein	*S. aurantiacus*	97%	436	WP_016639910.1
3631	sugar ABC transporter permease	*S. aurantiacus*	99%	323	WP_016639911.1
3632	carbohydrate ABC transporter permease	*S. aurantiacus*	99%	239	WP_016639912.1
3633	alpha-mannosidase	*Streptomyces* sp. NRRL B-1347	85%	1056	WP_030669426.1
	Cluster 20 T2 PKS				
3699	Hypo				
3700	xanthine dehydrogenase YagR molybdenum-binding subunit	*S. aurantiacus*	99%	758	WP_016640675.1
3701	xanthine dehydrogenase YagS FAD-binding subunit	*S. aurantiacus*	99%	334	WP_016640674.1
3702	xanthine dehydrogenase YagT iron-sulfur-binding subunit	*S. aurantiacus*	99%	182	WP_016640673.1
3703	two-component sensor histidine kinase VanS	*S. aurantiacus* *Streptomyces* sp. WAC1438	97% 82%	360	WP_037658942.1 AFS59919.1
3704	DNA-binding response regulator VanR	*S. aurantiacus* *Streptomyces* sp. WAC1438	98% 92%	231	WP_016640671.1 AFS59920.1
3705	peptidase M15	*Streptomyces* sp. MMG1533	86%	140	WP_078973198.1
3706	vancomycin resistance protein VanJ	*Micromonospora eburnea*	81%	379	SCL59618.1
3707	endoglycoceramidase	*S. aurantiacus*	96%	501	WP_016640668.1
3708	Hypo				
3709	NAD-dependent protein deacetylase SIR2 family transcriptional regulator	*S. aurantiacus* *S.* pristinaespiralis ATCC 25486	99% 84%	241	WP_016640666.1 EFH30723.1
3710	XRE family transcriptional regulator	*Streptomyces* sp. CB02009	49%	343	WP_073905763.1
3711	DUF4440 domain-containing protein EcaC	*S. aurantiacus* *S. stelliscabiei*	90% 83%	131	WP_078620987.1 KND42971.1
3712	arabinan endo-1,5-alpha-L-arabinosidase glycosyl hydrolase	*S. aurantiacus* *S. aureocirculatus*	94% 79%	338	WP_016640664.1 WP_030562017.1
3713	fructosamine kinase	*S. aurantiacus*	97%	293	WP_078621166.1
3714	Hypo				
3715	putative Tetracycline repressor protein class C TetR family transcriptional regulator	*S. aurantiacus* JA 4570 *S. aurantiacus*	99% 99%	223	EPH44400.1 WP_052029060.1
3716	MFS transporter	*S. lincolnensis*	87%	554	WP_067441517.1
3717	acyl carrier protein	*S. aurantiacus*	99%	76	WP_016640658.1
3718	3-oxoacyl-ACP synthase III	*S. aurantiacus*	99%	344	WP_016640657.1
3719	polyketide beta-ketoacyl synthase 1 NonK	*S. aurantiacus* *S. lincolnensis*	98% 48%	439	WP_016640656.1 ANS68650.1

3720	beta-ketoacyl synthase Ketoacyl-acyl carrier protein synthase II	*S. aurantiacus*	99% 90%	414	WP_063890344.1 AJS09384.1
3721	acetate CoA-transferase	*Streptomyces* sp. NRRL S-920	87%	640	WP_030784996.1
3722	3-oxoacyl-ACP synthase	*S. aurantiacus*	99%	319	WP_016640653.1
3723	3-oxoacyl-ACP synthase	*S. aurantiacus*	99%	344	WP_016640652.1
3724	3-oxoacyl-ACP reductase NAD(P)-dependent oxidoreductase	*S. aurantiacus* *Streptomyces* sp. NRRL S-920	99% 92%	256	WP_016640651.1 WP_037897386.1
3725	Ketoacyl-acyl carrier protein synthase III	*S.* alboniger	95%	336	AJS09389.1
3726	polyketide beta-ketoacyl synthase 1 NonK	*S. aurantiacus* *S. lincolnensis*	99% 86%	431	WP_016640649.1 ANS68650.1
3727	Ketoacyl-(acyl-carrier-protein) reductase	*S. alboniger*	85%	512	AJS09391.1
3728	NAD(P)-dependent oxidoreductase	*S. aurantiacus*	99%	275	WP_016640647.1
3729	ATP-dependent acyl-CoA ligase	*S. aurantiacus*	98%	572	WP_016645410.1
3730	Metal-dependent hydrolase (resistance)	*S. alboniger*	90%	327	AJS09394.1
3731	aspartate aminotransferase family protein	*S. phaeochromogenes*	85%	448	WP_055610996.1
3732	SAM-dependent methyltransferase putative Cypemycin methyltransferase	*S. aurantiacus* *S. aurantiacus* JA 4570	100% 100%	269	WP_052029430.1 EPH39634.1
3733	Enoyl-CoA hydratase	*S. alboniger*	87%	279	AJS09397.1
3734	DNA-binding response regulator	*Streptomyces* sp. NRRL S-920	91%	215	WP_051819621.1
3735	sterol 3-beta-glucosyltransferase UDP-glucose--sterol glucosyltransferase	*S. aurantiacus* *S. silvensis*	96% 79%	475	WP_016643542.1 WP_058847810.1
3736	spectinomycin resistance protein aminoglycoside O-phosphotransferase APH(9)-Ib	*S. spectabilis* *S. netropsis*	70% 66%	328	AAF63341.1 WP_063842186.1
	Cluster 21 Lantipeptide				
3782	ATPase AAA	*Streptomyces* sp. NRRL F-5727	76%	271	WP_031015667.1
3783	Hypo				
3784	Phage integrase, N-terminal SAM-like domain	*Streptomyces* sp. MnatMP-M17	89%	143	SCF64624.1
3785	integrase	*Streptomyces* sp. WM6373	74%	365	WP_053706923.1
3786	DNA-binding transcriptional regulator, XRE family	*Streptomyces* sp. MnatMP-M17	95%	104	SCF83829.1
3787	Tyrosine recombinase XerD	*Mycobacterium chlorophenolicum*	90%	116	KMO83348.1
3788	recombinase	*S. subrutilus*	93%	193	WP_069923487.1
3789	IstB domain-containing protein ATP-binding protein ATPase AAA	*S. azureus* *S. yerevanensis*	76% 77%	53	GAP50229.1 WP_033329515.1
3790	snsect kinin peptide	*S. azureus*	46%	293	GAP52679.1
3791	AraC family transcriptional regulator	*S. bingchenggensis*	84%	337	WP_043487438.1
3792	Dienelactone hydrolase and related enzymes	*Streptomyces* sp. SceaMP-e96	88%	298	SCK05969.1
3793	NUDIX domain-containing protein	*Streptomyces* sp. WM6378	67%	61	WP_053730618.1
3794	methyltransferase	*Streptomyces* sp. AA1529	71%	340	WP_037705968.1
3795	lanthionine synthetase	*Nocardiopsis dassonvillei*	62%	408	WP_061080312.1
3796	lantibiotic dehydratase	*Streptomyces* sp. AA1529	67%	1007	WP_026004409.1
3797	FxLD family lantipeptide	*S. yanglinensis*	69%	65	SEG85770.1
3798	protein-L-isoaspartate(D-aspartate) O-methyltransferase	*Streptomyces* sp. DI166	77%	409	SBT92801.1
3799	endonuclease/exonuclease/phosphatase family protein	*Streptomyces* sp. NBRC 110030	73%	301	WP_055469037.1
3800	Hypo				
3801	Hypo				
3802	ATP-binding protein	*S. pathocidini*	67%	144	WP_055473149.1

3803	Hypo				
3804	Hypo				
3805	Hypo				
3806					
3807	SARP family transcriptional regulator	*S. rimosus*	51%	759	KEF08288.1
	Cluster 22 Lantipeptide-T1-NRPS				
3847	Deoxyribonuclease NucA/NucB	*Streptomyces* sp. cf124	44%	345	SFM73502.1
3848	Hypo				
3849	oxidoreductase NAD(P)-dependent dehydrogenase, short-chain alcohol dehydrogenase family	*S. aurantiacus* *Saccharopolyspora hirsuta*	100% 80%	79	WP_016644894.1 SEG78731.1
3850	oxidoreductase	*S. aurantiacus*	97%	174	WP_016644894.1
3851	AfsR/SARP family transcriptional regulator	*S. aureocirculatus*	86%	265	WP_030565582.1
3852	AfsR/SARP family transcriptional regulator	*S. aureocirculatus*	92%	273	WP_030565582.1
3853	pentapeptide repeat protein	*Streptomyces* sp. NRRL S-813	64%	303	WP_030184811.1
3854	Hypo				
3855	chromosome partitioning protein ParB	*Lactobacillus paracasei* Cov 42	40%	119	WP_045137162.1
3856	Hypo				
3857	Site-specific DNA recombinase	*Streptomyces* sp. IgraMP-1	81%	188	SCE46538.1
3858	repressor LexA	Candidatus *Desulforudis audaxviator* Cov 68	45%	135	WP_012301867.1
3859	Hypo				
3860	Hypo				
3861	Hypo				
3862	transposase	*S. albulus*	92%	101	OAL12625.1
3863	predicted protein	*S. albus* J1074	92%	119	EFE82536.1
3864	Hypo				
3865	SMI1/KNR4 family protein	*S. aureocirculatus*	84%	189	WP_030573585.1
3866	HEAT repeat	*Streptomyces* sp. KS_16	81%	444	SDR58268.1
3867	Hypo				
3868	transcriptional regulator acyltransferase	*S. aurantiacus* *Streptomyces* sp. Sge12	97% 60%	534	WP_016638440.1 WP_081521097.1
3869	deaminase reductase	*S. purpureus*	92%	141	WP_019891614.1
3870	Hypo				
3871	putative Subtilin biosynthesis protein SpaB	*S. aurantiacus* JA 4570	99%	1086	EPH46638.1
3872	putative Subtilin biosynthesis protein SpaC	*S. aurantiacus* JA 4570	100%	434	EPH46637.1
3873	IS5 family transposase	*S. rimosus*	88%	141	WP_031008397.1
3874	resolvase	*Actinobacteria bacterium* 69-20	98%	249	OJV22912.1
3875	ArsR family transcriptional regulator	*S. vitaminophilus*	90%	239	WP_018381994.1
3876	flavin reductase NADH-FMN oxidoreductase RutF, flavin reductase (DIM6/NTAB) family	*Streptomyces* sp. Tu 6176 *S. yanglinensis*	92% 78%	231	WP_037894855.1 SEG52913.1
3877	Transposase and inactivated derivatives, TnpA family	*Streptomyces*	88%	142	SCK63098.1
3878	Hypo				
3879	integrase family protein (plasmid)	*S. pratensis* ATCC 33331	72%	779	ADW07849.1
3880	integrase	*Streptomyces* sp. SBT349	86%	357	WP_049571382.1
3881	Hypo				
3882	cytochrome P450	*Streptomyces* sp. MBT76	86%	396	WP_058043931.1
3883	Hypo				
3884					

3885	dihydrofolate reductase	*S. scabiei*	97%	178	WP_060904583.1
3886	Hypo				
3887	Hypo				
3888	putative Guanosine-3',5'-bis(diphosphate) 3'-pyrophosphohydrolase	*S. aurantiacus* JA 4570	90%	168	EPH46618.1
3889	calcium-binding protein	*Streptomyces* sp. H-KF8	92%	267	WP_065001537.1
3890	ATP-binding protein	*S. diastatochromogenes* Cov 70	59%	90	WP_060898477.1
3891	integrin-like protein	*Streptomyces* sp. 142MFCol3.1	88%	488	WP_037766030.1
3892	TetR family transcriptional regulator	*S. aurantiacus*	95%	205	WP_016638487.1
3893	Hypo				
3894	peptidoglycan-binding protein	*S. aurantiacus*	92%	241	WP_037657037.1
3895	DUF1838 domain containing protein	*S. fulvissimus* DSM 40593	87%	248	AGK78157.1
3896	putative Dimodular nonribosomal peptide synthase	*S. aurantiacus* JA 4570	95%	1079	EPH46610.1
3897	erythronolide synthase acyl transferase	*S. aurantiacus* *Streptomyces* sp. NRRL S-444	97% 82%	1040	WP_016638481.1 KJY47935.1
3898	polyketide synthase	*S. aurantiacus*	95%	2207	WP_016638480.1
3899	polyketide synthase	*S. aurantiacus*	94%	1899	WP_016638479.1
3900	putative Oleandomycin polyketide synthase, modules 5 and 6 Acyl transferase domain-containing protein	*S. aurantiacus* JA 4570 *S. yanglinensis*	97% 73%	1053	EPH46605.1 SEF53002.1
3901	thioesterase	*S. aurantiacus*	97%	245	WP_016638477.1
3902	cytochrome P450	*S. aurantiacus*	99%	416	WP_016638476.1
3903	Surfactin synthase subunit 1	*S. aurantiacus*	99%	89	WP_052028917.1
3904	daunorubicin ABC transporter permease	*S. aurantiacus*	99%	253	WP_016638474.1
3905	daunorubicin resistance protein DrrA family ABC transporter ATP-binding protein	*S. aurantiacus*	100%	312	WP_016638473.1
3906	thioesterase	*S. aurantiacus*	97%	241	WP_016638472.1
3907	amino acid adenylation domain-containing protein	*S. yanglinensis*	81%	1274	SEF52767.1
3908	non-ribosomal peptide synthase domain TIGR01720/amino acid adenylation domain-containing protein	*S. yanglinensis*	75%	2574	SEF52734.1
3909	non-ribosomal peptide synthase domain TIGR01720/amino acid adenylation domain-containing protein	*S. yanglinensis*	78%	2882	SEF52696.1
3910	NAD(P)/FAD-dependent oxidoreductase	*S. aurantiacus*	100%	394	WP_016638468.1
3911	phosphopantetheine-binding protein	*S. aurantiacus*	97%	76	WP_037657033.1
3912	peptide synthetase	*S. mirabilis*	82%	530	WP_037751281.1
3913	regulator	*S. cellostaticus*	76%	240	WP_067003863.1
3914	MbtH family protein	*S. aurantiacus*	99%	71	WP_016638464.1
3915	putative 3-methyl-2-oxobutanoate dehydrogenase subunit alpha pyruvate dehydrogenase	*S. aurantiacus* *S. phaeoluteigriseus*	98% 89%	279	WP_016638463.1 WP_073491954.1
3916	alpha-ketoacid dehydrogenase subunit beta	*S. griseus*	89%	326	WP_050513872.1
3917	dihydrolipoamide succinyltransferase	*S. aurantiacus*	100%	75	WP_037657045.1
3818	Acyl carrier protein	*S. yanglinensis*	90%	82	SEF52432.1
3819	ketoacyl-ACP synthase III	*S. aurantiacus*	99%	312	WP_016638459.1
3820	zinc-binding dehydrogenase NADPH:quinone reductase	*S. phaeoluteigriseus* *S. yanglinensis*	87% 80%	378	OQD51826.1 SEF52379.1
3821	cytochrome P450	*S. aurantiacus*	97%	391	WP_016638457.1
3822	flavin reductase	*S. aurantiacus*	97%	182	WP_016638456.1
3823	polyketide synthase Acyl carrier protein	*S. aurantiacus* *Streptomyces* sp. DvalAA-43	98% 47%	91	WP_037657028.1 SCD55195.1
3824	MFS transporter	*S. cyaneogriseus*	89%	460	WP_044378172.1
3825	acyl-CoA desaturase	*S. aurantiacus*	99%	368	WP_016641448.1

	fatty acid desaturase	*S. ghanaensis* ATCC 14672	93%		EFE72574.1
3826	Hypo				
3827	putative Lysylphosphatidylglycerol biosynthesis bifunctional protein LysX VlmA	*S. aurantiacus* *S. ghanaensis*	92% 85%	368	WP_016641446.1 WP_004994381.1
3828	diacylglycerol kinase	*S. leeuwenhoekii*	96%	357	WP_016641445.1
3829	VlmK-like protein	*S. cyaneogriseus*	91%	489	WP_044378185.1
3830	putative Lactaldehyde dehydrogenase	*S. aurantiacus*	99%	535	WP_016641443.1
3831	acyl-CoA dehydrogenase	*S. aurantiacus*	99%	379	WP_037659864.1
3832	aspartate aminotransferase family protein	*S. aurantiacus*	99%	428	WP_016641441.1
3833	cyclase	*S. viridosporus*	81%	161	WP_016828574.1
	Cluster 23 T1PKS-NRPS				
3915	putative 3-methyl-2-oxobutanoate dehydrogenase subunit alpha	*S. aurantiacus*	98%	279	WP_016638463.1
3916	pyruvate dehydrogenase E1 component subunit beta	*S. aurantiacus*	99%	326	WP_016638462.1
3917	dihydrolipoamide succinyltransferase	*S. aurantiacus*	100%	75	WP_037657045.1
3918	Acyl carrier protein	*S. yanglinensis*	90%	82	SEF52432.1
3919	ketoacyl-ACP synthase III	*S. aurantiacus*	99%	312	WP_016638459.1
3920	zinc-binding dehydrogenase	*S. phaeoluteigriseus*	87%	378	OQD51826.1
3921	cytochrome P450	*S. aurantiacus*	97%	391	WP_016638457.1
3922	flavin reductase		97%	182	WP_016638456.1
3923	polyketide synthase	*S. aurantiacus*	98%	91	WP_037657028.1
3924	MFS transporter	*S. cyaneogriseus*	89%	460	WP_044378172.1
3925	acyl-CoA desaturase	*S. aurantiacus*	99%	368	WP_016641448.1
3926	H				
3927	putative Lysylphosphatidylglycerol biosynthesis bifunctional protein LysX VlmA	*S. aurantiacus* *S. ghanaensis*	92% 85%	368	WP_016641446.1 WP_004994381.1
3928	diacylglycerol kinase	*S. leeuwenhoekii*	76%	357	WP_029385951.1
3929	VlmK-like protein	*S. cyaneogriseus*	91%	489	WP_044378185.1
3930	aldehyde dehydrogenase Acyl-CoA reductase	*S. leeuwenhoekii* *S. guanduensis*	95% 76%	535	WP_029385949.1 SDO37109.1
3931	acyl-CoA dehydrogenase	*S. aurantiacus*	99%	379	WP_037659864.1
3932	aspartate aminotransferase family protein	*S. aurantiacus*	99%	428	WP_016641441.1
3933	cyclase	*S. viridosporus*	96%	161	WP_016641440.1
3934	Hypo				
3935	VlmB-like protein	*S. aurantiacus*	99%	333	WP_037662717.1
3936	ketoacyl-ACP synthase III TmcD	*S. aurantiacus* *S. chromofuscus*	97% 71%	339	WP_016644041.1 CUX96951.1
3937	acyl carrier protein	*S. aurantiacus*	99%	77	WP_016644042.1
3938	serine--tRNA ligase	*S. aurantiacus*	98%	424	WP_016645372.1
3939	SARP family transcriptional regulator	*Streptomyces* sp. CNQ431	82%	283	WP_033947309.1
3940	nuclear transport factor 2 family protein	*S. aurantiacus*	99%	123	WP_016645370.1
3941	MFS transporter tetracenomycin C resistance and export protein	*S. aurantiacus* *S. ghanaensis* ATCC 14672	99% 84%	485	WP_016645369.1 EFE72590.1
3942	acyl carrier protein	*S. cyaneogriseus*	82%	107	WP_044387231.1
3943	type I polyketide synthase	*S. aurantiacus*	92%	2059	WP_016645367.1
3944	4-coumarate--CoA ligase	*S. aurantiacus*	97%	538	WP_016645366.1
3945	monooxygenase LLM class flavin-dependent oxidoreductase	*S. aurantiacus* *S. silaceus*	98% 72%	328	WP_037664244.1 WP_076684750.1
3946	thioesterase	*S. aurantiacus*	99%	264	WP_016645364.1
3947	alpha/beta hydrolase	*S. aurantiacus*	97%	276	WP_037664243.1
3948	Hypo				
3949	4-phosphopantetheinyl transferase	*S. aurantiacus*	97%	315	WP_052029425.1
3950	Hypo				
3951	acyl-CoA dehydrogenase	*S. aurantiacus*	98%	594	WP_037664241.1
3952	thioesterase	*S. aurantiacus*	98%	261	WP_016645359.1
3953	putative transcriptional regulator RedD	*S. aurantiacus* JA 4570	82%	373	EPH39676.1

3954	peptide synthase condensation domain-containing protein	*S. leeuwenhoekii*	67%	818	WP_029386407.1
3955	Granaticin polyketide synthase bifunctional cyclase/dehydratase	*S. leeuwenhoekii*	80%	319	CQR59803.1
3956	Regulator protein	*S. aurantiacus*	99%	273	WP_016645354.1
3957	DNA-binding transcriptional activator of the SARP family	*Streptomyces* sp. MnatMP-M17	76%	284	SCF79391.1
3958	putative A-factor receptor protein	*S. aurantiacus*	99%	218	WP_016643811.1
3959	methylmalonyl-CoA carboxyltransferase	*Streptomyces* sp. NRRL B-1347	90%	566	WP_030676109.1
3960	Acyl-CoA carboxylase epsilon subunit	*Streptomyces* sp. 2131.1	48%	89	SEE80268.1
3961	DNA-binding transcriptional activator of the SARP family	*S.* sp. MnatMP-M17	67%	362	SCF79391.1
3962	cytochrome P450	*S. aurantiacus*	95%	404	WP_078621484.1
3963	TetR/AcrR family transcriptional regulator	*S. aurantiacus*	99%	228	WP_037662025.1
3964	acyl-CoA oxidase	*S. glaucescens* Cov 52	55%	1127	WP_078957887.1
3965	Low				
3966	Low				
3967	Low				
3968	Low				
3869	Low				
3870	VWA domain-containing protein	*Streptomyces* sp. RV15	97%	352	WP_067021145.1
	Cluster 24 NRPS-transat PKS-Other KS				
3964	acyl-CoA oxidase	*S. glaucescens* Cov 52	55%	1127	WP_078957887.1
3965	Low				
3966					
3967	Hypo				
3968	Hypo				
3969	Hypo				
3970	VWA domain-containing protein	*Streptomyces* sp. RV15	97%	352	WP_067021145.1
3971	No				
3972	DinB family protein	*S. aurantiacus*	97%	199	WP_016641195.1
3973	succinate dehydrogenase flavoprotein subunit	*S. laurentii*	67%	89	BAU81599.1
3974	Low				
3975					
3976	NAD-dependent glycerol-3-phosphate dehydrogenase	*Streptomyces* sp. W007	72%	81	EHM30912.1
3977	Peptidoglycan-binding domain 1 protein	*Cellulomonas flavigena* DSM 20109	43%	52	ADG76564.1
3978	Leu-binding protein	*S. aurantiacus*	96%	370	WP_016641199.1
3979	pentapeptide repeat protein	*S. aurantiacus*	99%	275	WP_016641200.1
3980	cytochrome P450	*S. aurantiacus*	99%	410	WP_037659573.1
3981	Hypo				
3982	Hypo				
3983	acyl-carrier-protein S-malonyltransferase	*S. aurantiacus*	97%	646	WP_016641205.1
3984	putative Polyketide biosynthesis protein PksE	*S. aurantiacus*	97%	805	WP_016641206.1
3985	Polyketide biosynthesis 3-hydroxy-3-methylglutaryl-ACP synthase PksG	*S. aurantiacus*	99%	421	WP_016641207.1
3986	Polyketide biosynthesis malonyl-ACP decarboxylase PksF BryQ	*S. aurantiacus* *Candidatus Endobugula sertula*	99% 60%	707	WP_016641208.1 ABK51298.1
3987	Polyketide biosynthesis acyl-carrier-protein AcpK	*S. aurantiacus* Cov 81	100%	103	WP_016641209.1
3988	non-ribosomal peptide synthetase	*Streptomyces* sp. DSM 15324	85%	3117	WP_079079242.1
3989	non-ribosomal peptide synthetase	*Streptomyces* sp. ERV7	72%	2776	WP_067167615.1
3990	putative Polyketide synthase PksL	*S. aurantiacus* JA 4570	92%	4769	EPH43450.1
3991	trans-AT hybrid polyketide synthase-	*Nostoc* sp.	36%	7414	ADA69239.2

	NRPS				
3992	mixed polyketide synthase/non-ribosomal peptide synthetase Malonyl CoA-acyl carrier protein transacylase	*Saccharomonospora azurea* SZMC 14600 *Bacillus amyloliquefaciens*	42% 40%	4982	EHK80196.1 KYC88114.1
3993	thioesterase *S. aurantiacus* oleoyl-ACP hydrolase	*S. aurantiacus* *Streptomyces* sp. SAT1	98% 64%	251	WP_016643751.1 ANO42184.1
3994	putative 6-aminohexanoate-cyclic-dimer hydrolase	*S. aurantiacus* *Streptomyces* sp. NRRL B-1347	99%	460	WP_016643750.1 WP_078868529.1
3995	lipase	*Streptomyces* sp. 13-12-16	88%	67	WP_085576446.1
3996	cupin	*S. aurantiacus*	97%	155	WP_037662345.1
3997	AraC family transcriptional regulator	*S. aurantiacus*	97%	262	WP_037662343.1
3998	Hypo				
3999	TetR/AcrR family transcriptional regulator	*S. aurantiacus*	99%	210	WP_016643745.1
4000	MFS transporter	*S. aurantiacus*	94%	402	WP_016643744.1
4001	NAD(P)-dependent oxidoreductase	*S. niger*	88%	197	WP_052865511.1
4002	Hypo				
4003	carboxylesterase/lipase family protein	*S. aurantiacus*	97%	482	WP_037662305.1
4004	MerR family transcriptional regulator	*S. silvensis*	74%	137	WP_058846801.1
4005	Hypo				
4006	Hypo				
4007	TetR family transcriptional regulator	*S. aurantiacus*	97%	189	WP_016643737.1
4008	amidase	*S. aurantiacus*	94%	562	WP_016643736.1
4009	AfsR family transcriptional regulator	*S. cattleya*	57%	647	WP_078590713.1
4010	MerR family transcriptional regulator	*Streptomyces* sp. NBRC 110027	88%	319	WP_042158732.1
4011	Hypo				
4012	daunorubicin resistance protein DrrA family ABC transporter ATP-binding protein	*Lechevalieria aerocolonigenes* Cov 56	74%	60	WP_030468215.1
4013	GH3 auxin-responsive promoter	*Streptomyces* sp. cf386	50%	559	SDP81041.1
	Cluster 25 NRPS				
4078	Hypo				
4079	ATPase AAA	*S. bingchenggensis*	74%	363	WP_014173264.1
4080	tetratricopeptide repeat protein	*S. aurantiacus*	99%	1490	WP_078621013.1
4081	HEXXH motif domain-containing protein		95%	637	WP_016639666.1
4082	Hypo				
4083	Hypo				
4084	branched-chain amino acid ABC transporter substrate-binding protein	*S. aurantiacus*	99%	919	WP_016639669.1
4085	cupin	*Streptomyces* sp. NRRL F-4489	84%	147	WP_066982294.1
4086	ATP-dependent DNA ligase	*Streptomyces* sp. NRRL B-1347	72%	318	WP_030688459.1
4087	beta-phosphoglucomutase family hydrolase	*Streptomyces* sp. HGB0020	79%	261	WP_016430638.1
4088	glycosyl hydrolase	*S. aurantiacus*	99%	781	WP_016639673.1
4089	Hypo				
4090	N-acetyltransferase	*S. aurantiacus*	97%	170	WP_016639676.1
4091	antibiotic synthesis protein MbtH	*S. pharetrae*	94%	67	WP_086171982.1
4092	putative Linear gramicidin synthase subunit C	*S. aurantiacus* JA 4570	97%	6769	EPH45403.1
4093	non-ribosomal peptide synthetase	*S. silvensis*	87%	6254	WP_058846934.1
4094	penicillin-binding protein	*S. aurantiacus*	99%	449	WP_037658623.1
4095	cytochrome P450	*S. aurantiacus*	99%	404	WP_016640331.1
4096	chaplin	*S. aurantiacus*	95%	79	WP_016640333.1
4097	dihydrodipitnate reductase	*S. atriruber*	84%	359	WP_055567218.1
4098	carboxymuconolactone decarboxylase family protein	*S. aurantiacus*	99%	163	WP_016640335.1

4099	Hypo				
4100	Hypo				
4101	argininosuccinate synthase	*S. aurantiacus*	100%	481	WP_016640338.1
4102	LysR family transcriptional regulator	*Actinosporangium* sp. NRRL B-3428	76%	288	WP_052575516.1
4103	putative protein PecM	*S. aurantiacus*	98%	320	WP_016640341.1
4104	NADH-flavin reductase	*Actinosporangium* sp. NRRL B-3428	69%	212	WP_052575518.1
4105	RNA polymerase sigma-F factor	*S. aurantiacus*	94%	319	WP_016640343.1
4106	serine/threonine protein phosphatase	*S. aurantiacus*	99%	399	WP_037658638.1
4107	lactate 2-monooxygenase	*S. aurantiacus*	97%	388	WP_016640345.1
4108	ABC transporter permease	*S. aurantiacus*	99%	403	WP_016640346.1
4109	Macrolide export ATP-binding/permease MacB	*S. aurantiacus*	96%	313	WP_016640347.1
4110	Hypo				
4011	Hypo				
4012	DNA-binding response regulator	*S. aurantiacus*	99%	217	WP_016640350.1
	Cluster 26 Other				
4123	GNAT family N-acetyltransferase	*S. aurantiacus*	98%	190	WP_016640361.1
4124	N-acetyltransferase	*S. aurantiacus*	96%	215	WP_078621133.1
4125	sphingomyelin phosphodiesterase	*S. aurantiacus*	98%	326	WP_016640362.1
4126	MarR family transcriptional regulator	*S. humi*	64%	215	WP_046729482.1
4127	EamA/RhaT family transporter	*Streptomyces* sp. CNS606	68%	298	WP_027764799.1
4128	putative Aminoglycoside 6-adenylyltransferase	*S. aurantiacus*	98%	280	WP_016640366.1
4129	putative Carotenoid cleavage oxygenase	*S. aurantiacus* JA 4570	96%	514	EPH44680.1
4130	transcriptional regulator *S. aurantiacus*	*S. aurantiacus*	89%	383	WP_052029045.1
4131	glyoxalase/bleomycin resistance/extradiol dioxygenase family protein	*Microtetraspora fusca*	87%	141	WP_066942264.1
4132	winged helix DNA-binding domain-containing protein	*S. aurantiacus*	98%	368	WP_037658687.1
4133	carbamoyltransferase	*Streptomyces* sp. PRh5	89%	593	WP_037955810.1
4134	putative xanthine dehydrogenase YagT iron-sulfur-binding subunit	*S. aurantiacus*	90%	219	WP_016640371.1
4135	xanthine dehydrogenase YagS FAD-binding subunit molybdopterin dehydrogenase	*S. aurantiacus* *S. silvensis*	99% 91%	330	WP_016640372.1 WP_058847890.1
4136	putative xanthine dehydrogenase YagR molybdenum-binding subunit	*S. aurantiacus* JA 4570	98%	700	EPH44686.1
4137	Hypo				
4138	Cupin domain protein *S. melanosporofaciens*		71%	144	SED63004.1
4139	histidinol-phosphate aminotransferase	*S. aurantiacus*	96%	653	WP_016640376.1
4140	NAD(P)-dependent oxidoreductase	*S. aurantiacus*	97%	315	WP_016640377.1
4141	polyketide synthase peptide synthetase Acyl-CoA synthetase (AMP-forming)/AMP-acid ligase II	*S. aurantiacus* *S. bingchenggensis* *S. melanosporofaciens*	96% 80% 80%	926	WP_016640378.1 WP_014181622.1 SED62952.1
4142	argininosuccinate lyase 2	*S. aurantiacus*	95%	386	WP_037658647.1
4143	argininosuccinate lyase 2	*S. aurantiacus*	97%	388	WP_037658648.1
4144	SGNH/GDSL hydrolase family protein	*Actinobacteria bacterium* OV450	84%	321	WP_054223314.1
4145	transcriptional repressor	*S. aurantiacus*	99%	136	WP_016640381.1
4146	catalase-peroxidase	*S. aurantiacus*	99%	741	WP_016640382.1
4147	calcium-binding protein	*S.* curacoi	92%	84	WP_062152579.1
4148	endoglucanase	*S. aurantiacus*	98%	352	WP_016640384.1
4149	Hypo				
4150	phage tail protein	*S. sparsogenes*	86%	190	WP_065967665.1
4151	Hypo				
4152	Hypo				
4153					

4154	Hypo				
4155	Hypo				
4156	Hypo				
4157	Hypo				
4158	Hypo				
4159	aminoglycoside 3-N-acetyltransferase	*S. aurantiacus*	98%	282	WP_016640395.1
4160	glutaminase	*S. aurantiacus*	99%	308	WP_016640396.1
4161	Hypo				
4162	RidA family protein	*S. aurantiacus*	99%	135	WP_016640398.1
4163	transcriptional regulator	*S. aurantiacus*	100%	126	WP_016640399.1
4164	GNAT family N-acetyltransferase	*S. aurantiacus*	97%	340	WP_016640400.1
	Cluster 27 Terpene				
4170	NAD(+) synthase	*S. aurantiacus*	98%	692	WP_016640406.1
4171	gluconolactonase	*S. aurantiacus*	99%	319	WP_016640407.1
4172	Lrp/AsnC family transcriptional regulator	*S. aurantiacus*	99%	150	WP_016640408.1
4173	putative RdlA protein	*Streptomyces* sp. Tu6071	75%	134	EGJ75635.1
4174	Hypo				
4175	polyphosphate kinase 2	*S. aurantiacus*	99%	324	WP_052029042.1
4176	Low				
4177	magnesium transporter MgtC	*S. silvensis*	85%	243	WP_058847096.1
4178	Hypo				
4179	glycoprotein gp2	*S. aurantiacus*	96%	296	WP_016640415.1
4180	MxaD family protein	*S. aurantiacus*	96%	160	WP_016640416.1
4181	Antibiotic biosynthesis monooxygenase	*Streptomyces* sp. SceaMP-e96	50%	239	SCK22898.1
4182	geranyl diphosphate 2-C-methyltransferase SAM-dependent methyltransferase	*S. aurantiacus* *S. aureocirculatus*	96% 81%	312	WP_016640418.1 WP_030558778.1
4183	cyclase Chain A, Crystal Structure Of A Terpene Synthase	*S. aurantiacus* *S. Lydicus*, Target Efi-540129	98% 78%	353	WP_037658662.1 4ZQ8_A
4184	protein SrpI Crp/Fnr family transcriptional regulator	*S. aurantiacus* *S.* mirabilis	99% 82%	479	WP_016640420.1 WP_037740841.1
4185	radical SAM protein	*S. aurantiacus*	99%	376	WP_016640421.1
4186	ABC transporter ATP-binding protein	*S. aurantiacus*	99%	600	WP_037658663.1
4187	putative Iron-sulfur clusters transporter ATM1	*S. aurantiacus* JA 4570	95%	700	EPH44736.1
4188	Low				
	Cluster 28 NRPS				
4232	Hypo				
4233	GntR family transcriptional regulator	*S. aurantiacus*	99%	227	WP_016643911.1
4234	putative D-3-phosphoglycerate dehydrogenase	*S. aurantiacus* JA 4570	98%	336	EPH41137.1
4235	putative hydroxypyruvate isomerase	*S. aurantiacus*	96%	295	WP_016643909.1
4236	Asp/Glu racemase	*Stackebrandtia nassauensis* Cov 84	49%	233	WP_013016830.1
4237	Hrp-dependent type III effector protein	*S. kanamyceticus*	83%	436	WP_055553261.1
4238	molybdenum ABC transporter substrate-binding protein	*S. anulatus* Cov 50	64%	464	WP_057665579.1
4239	C4-dicarboxylate ABC transporter permease	*S. aureocirculatus*	94%	449	WP_030558847.1
4240	Hypo				
4241	tetratricopeptide repeat-containing protein	*S. aurantiacus*	97%	459	WP_016643904.1
4242	Hypo				
4243	Hypo				
4244	integral membrane protein	*Streptomyces* sp. C	82%	88	EFL12840.1
4245	putative F420-dependent glucose-6-phosphate dehydrogenase	*S. aurantiacus* JA 4570	99%	323	EPH41126.1

4246	cyclase	*S. aureocirculatus*	86%	154	WP_030558854.1
4247	twin-arginine translocation pathway signal protein	*S. albus* subsp. albus	67%	152	KUJ70222.1
4248	membrane protein	*S.* sp. NRRL WC-3753	68%	249	KPC71734.1
4249	Hypo				
4250	putative transcriptional activator FeaR araC family regulatory protein	*S. aurantiacus* *S. lividans* 1326	97% 62%	333	WP_016643892.1 EOY50001.1
4251	Peptidase inhibitor family I36	*Streptomyces* sp. MnatMP-M27	45%	150	SCG12871.1
4252	nuclear transport factor 2 family protein	*S. acidiscabies*	44%	151	WP_010353406.1
4253	FAD-binding monooxygenase putative tryptophan hydroxylase VioD	*S. aurantiacus* JA 4570	98% 98%	407	WP_037662522.1 EPH41117.1
4254	N-acetyltransferase	*S. aurantiacus*	87%	186	WP_016643889.1
4255	putative Dimodular nonribosomal peptide synthase	*S. aurantiacus*	97%	1354	WP_016643888.1
4256	putative Dimodular nonribosomal peptide synthase	*S. aurantiacus* JA 4570	95%	3166	EPH41114.1
4257	GSCFA family protein	*S. varsoviensis*	78%	349	KOG87483.1
4258	Hypo				
4259	alpha-ketoglutarate-dependent taurine dioxygenase similar to dioxygenase	*S. aurantiacus* *S. spectabilis*	99% 87%	290	WP_016643883.1 AAD50457.1
4260	SpcZ	*S. spectabilis*	67%	279	AAD50456.1
4261	DUF2848 domain-containing protein NanT2	*S. phaeochromogenes* *S. nanchangensis*	88% 87%	228	WP_055616689.1 AAP42849.1
4262	MFS transporter NanT1	*S. bingchenggensis* *S. nanchangensis*	85% 87%	466	WP_014180996.1 AAP42848.1
4263	acyltransferase	*S. sparsogenes*	64%	275	WP_065967225.1
4264	GntR family transcriptional regulator	*S. aurantiacus*	97%	208	WP_016643878.1
4265	NADP-dependent succinic semialdehyde dehydrogenase	*S. aurantiacus*	99%	472	WP_037662544.1
4266	PHP domain-containing protein histidinol phosphatase	*S. aurantiacus* *Streptomyces* sp. TP-A0875	99% 83%	346	WP_016643877.1 WP_078962515.1
4267	serine protease streptogrisin D/glutamyl endopeptidase II	*S. silvensis* *Streptomyces* sp. ScaeMP-e83	87% 81%	354	WP_058850947.1 SCE08117.1
4268	amidase	*S. aurantiacus*	95%	530	WP_037662519.1
4269	CBS domain-containing protein histidine kinase	*S. aurantiacus* *Streptomyces* sp. NRRL S-920	100% 81%	138	WP_016642111.1 WP_078594014.1
4270	iron-containing redox enzyme family protein	*S. aurantiacus*	97%	336	WP_016642112.1
4271	CDGSH iron-sulfur domain-containing protein	*S. aurantiacus*	96%	82	WP_037660533.1
4272	methyltransferase	*S. aurantiacus*	98%	238	WP_052029163.1
4273	Biotin carboxylase	*S.* wuyuanensis	47%	434	SDM06335.1
4274	putative membrane protein	*S.* avermitilis MA-4680 = NBRC 14893	71%	51	BAC70068.1
4275	alkaline D-peptidase putative D-alanyl-D-alanine carboxypeptidase	*S. aurantiacus* *S. aurantiacus* JA 4570	98% 97%	393	WP_078621338.1 EPH42953.1
4276	cytochrome P450	*S. aurantiacus*	97%	399	WP_016642120.1
	Cluster 29 Bacteriocin-TPKS-NRPS				
4329	PucR C-terminal helix-turn-helix domain-containing protein	*Streptomyces* sp. 2114.2	52%	702	SDT79859.1
4330	transcriptional regulator	*S. aurantiacus*	98%	281	WP_037662942.1
4331	S-adenosyl methyltransferase	*S. silvensis*	92%	264	WP_058847680.1
4332	PucR C-terminal helix-turn-helix domain-containing protein	*S.* sp. MnatMP-M77	47%	495	SBU99676.1
4333	membrane protein	*S. aurantiacus*	98%	745	WP_016643205.1
4334	putative zinc metalloprotease	*S. aurantiacus*	96%	247	WP_016643208.1
4335	DNA ligase-like protein 3'-phosphoesterase	*S. aurantiacus* *Streptomyces* sp.	95% 74%	224	WP_016643209.1 WP_078564337.1

		CNQ329			
4336	putative Glycine amidinotransferase	*S. aurantiacus* JA 4570	99%	373	EPH41825.1
4337	Hemerythrin HHE cation binding domain-containing protein	*Streptomyces* sp. 2131.1	80%	186	SEE37526.1
4338	NAD(P)-dependent oxidoreductase	*S. aurantiacus*	99%	263	WP_037661779.1
4339	thiamine pyrophosphate-requiring protein	*S. aurantiacus*	99%	616	WP_016643218.1
4340	methylesterase 19 alpha/beta hydrolase	*S. aurantiacus* *Streptomyces* sp. WM6372	97% 91%	287	WP_016643220.1 WP_053690235.1
4341	GNAT family N-acetyltransferase	*Actinobacteria bacterium* OK074	87%	157	WP_082414929.1
4342	transcriptional regulator	*S. aurantiacus*	98%	131	WP_016643221.1
4343	NmrA family transcriptional regulator putative NAD(P)H azoreductase	*S. aurantiacus* *S. aurantiacus* JA 4570	96% 96%	281	WP_037661785.1 EPH41836.1
4344	threonine/serine dehydratase	*S. aurantiacus*	95%	345	WP_016643223.1
4345	TetR family transcriptional regulator	*S.* misionensis	83%	230	WP_074990036.1
4346	Hypo				
4347	RiPP maturation radical SAM protein 1 ribosomal peptide maturation radical SAM protein 1	*S. aurantiacus* *Streptomyces* sp. SolWspMP-5a-2	98% 70%	640	WP_016643226.1 SCD94229.1
4348	peptidase S9	*Streptomyces* sp. Tu 6176	48%	34	EYT82211.1
4349	HAD family phosphatase	*S. ghanaensis*	88%	203	WP_050793842.1
4350	transcriptional regulator	*S. aurantiacus*	99%	150	WP_016643229.1
4351	putative transporter YgaY	*S. aurantiacus* JA 4570	98%	417	EPH41844.1
4352	electron transporter	*Streptomyces* sp. ERV7	80%	180	WP_067161958.1
4353	TetR family transcriptional regulator DNA-binding transcriptional regulator, AcrR family	*S. aurantiacus* *Jiangella* sp. DSM 45060	97% 74%	221	WP_052029241.1 SDS87312.1
4354	Hypo				
4355	Hypo				
4356	serine protease pkd repeat-containing protein	*S. aurantiacus* *S. aureocirculatus*	97% 84%	392	WP_016643237.1 WP_030559402.1
4357	proline racemase	Saccharopolyspora *shandongensis*	69%	341	SDZ14208.1
4358	Hypo				
4359	putative Linear gramicidin synthase subunit C NRPS	*S. aurantiacus* JA 4570	97%	3794	EPH41854.1
4360	3-hydroxybutyryl-CoA dehydrogenase	*S. aurantiacus*	97%	349	WP_052029247.1
4361	haloacid dehalogenase FkbH	*Rhodococcus jostii* *Rhodococcus opacus*	67% 67%	360	WP_011595010.1 AII08972.1
4362	acyl-CoA dehydrogenase	*S. aurantiacus*	97%	392	WP_078621468.1
4363	acyl carrier protein	*Streptomyces* sp. NRRL B-1347	77%	94	WP_030669506.1
4364	3-oxoacyl-ACP synthase	*Streptomyces* sp. NRRL B-1347	83%	335	WP_030669503.1
4365	non-ribosomal peptide synthetase Linear gramicidin synthase subunit C	*Streptomyces* sp. NRRL B-1347 *Rhodococcus opacus* PD630	72% 43%	5139	WP_051854505.1 AHK31361.1
4366	LuxR family transcriptional regulator	*S. ghanaensis*	87%	257	WP_004978547.1
4367	amidohydrolase	*S. aurantiacus*	97%	570	WP_016642019.1
4368	DNA-binding transcriptional regulator, HxlR family	*Nonomuraea solani*	84%	161	SEH02978.1
4369	short-chain dehydrogenase	*S. aurantiacus*	98%	233	WP_037660391.1
4370	Phenylpyruvate tautomerase PptA, 4-oxalocrotonate tautomerase family	*Nonomuraea solani*	85%	141	SEH02976.1
4371	ANTAR domain-containing protein	*S. aurantiacus*	98%	131	WP_016642015.1
4372	Hypo				
4373	phage holin family protein	*S. aurantiacus*	98%	144	WP_037660413.1
4374	YihY/virulence factor BrkB family protein	*S. aurantiacus*	95%	361	WP_016642012.1
4375	Putative peptidoglycan binding domain-containing protein	*Thermomonospora chromogena*	41%	333	SDQ59793.1

4376	guanylate cyclase	*Streptomyces* sp. NRRL S-920	91%	82	WP_030778038.1
4377	Metal-dependent hydrolase, endonuclease/exonuclease/phosphatase family	*Streptomyces* sp. cf386	81%	296	SDN56422.1
4378	Hypo				
4379	transcriptional regulator	*Streptomyces* sp. TOR3209	81%	135	WP_019328868.1
4380	serine/arginine repetitive matrix protein 2	*Streptomyces* sp. SPB074	45%	85	EDY43927.1
4381	ornithine carbamoyltransferase	*S. aurantiacus*	96%	314	WP_037660389.1
4382	non-ribosomal peptide synthetase	*S. aurantiacus*	98%	2568	WP_016642004.1
4383	non-ribosomal peptide synthetase	*S. aurantiacus*	94%	2215	WP_016642003.1
4384	thioesterase	*S. aurantiacus*	99%	257	WP_016642002.1
4385	Low				
4386	non-ribosomal peptide synthetase hybrid non-ribosomal peptide synthetase/type I polyketide synthase, partial	*S. aurantiacus* *Archangium violaceum*	98% 43%	2151	WP_016642001.1 WP_043410518.1
4387	putative Linear gramicidin synthase subunit D non-ribosomal peptide synthase	*S. aurantiacus* JA 4570 *Modestobacter* sp. Leaf380 Cov 43	99% 47%	1109	EPH43045.1 WP_056291109.1
4388	non-ribosomal peptide synthetase	*S. aurantiacus*	98%	1712	WP_016639311.1
4389	monooxygenase	*S. aurantiacus*	98%	323	WP_016639310.1
4390	antibiotic synthesis protein MbtH	*Modestobacter* sp. Leaf380	78%	67	WP_056301182.1
4391	putative Phthiocerol/phenolphthiocerol synthesis polyketide synthase type I PpsE	*S. aurantiacus* JA 4570	98%	1854	EPH45771.1
4392	MFS transporter	*S. aurantiacus*	98%	430	WP_016639306.1
4393	non-heme bromoperoxidase BpoC	*S. aurantiacus*	98%	268	WP_016639305.1
4394	DNA-binding response regulator	*S. aurantiacus*	100%	212	WP_016639304.1
4395	two-component sensor histidine kinase	*S. aurantiacus*	98%	414	WP_016639303.1
4396	ABC transporter permease	*S. aurantiacus*	99%	280	WP_016639302.1
4397	ABC transporter ATP-binding protein	*S. aurantiacus*	99%	297	WP_037657753.1
4398	putative siderophore biosynthesis protein SbnA	*S. aurantiacus* JA 4570	97%	340	EPH45763.1
4399	aspartate 1-decarboxylase	*S. aurantiacus*	98%	340	WP_052028969.1
4400	Hypo				
4401	putative Tripeptidyl aminopeptidase	*S. aurantiacus* JA 4570	98%	552	EPH45760.1
4402	ornithine cyclodeaminase	*S. aurantiacus*	99%	357	WP_016639296.1
4403	putative Regulator protein	*S. aurantiacus* JA 4570	98%	434	EPH45758.1
4404	Hypo				
4405	citrate synthase/methylcitrate synthase	*Streptomyces* sp. ScaeMP-e10	87%	382	WP_018514775.1
4406	methylisocitrate lyase	*Streptomyces* sp. TP-A0874	90%	301	WP_069813196.1
4407	2-methylcitrate dehydratase MmgE/PrpD family protein	*S. alni* *Streptomyces* sp. TP-A0874	88% 86%	507	SFE48074.1 WP_069813198.1
4408	XRE family transcriptional regulator	*S. atriruber*	89%	501	WP_055564255.1
4409	XRE family transcriptional regulator	*Streptomyces* sp. WM6378	91%	472	WP_053728065.1
	Cluster 30 NRPS				
4522	ABC transporter permease	*S. aurantiacus*	98%	357	WP_016639403.1
4523	ABC transporter permease	*S. aurantiacus*	99%	337	WP_016639402.1
4524	sugar ABC transporter ATP-binding protein	*S. aurantiacus*	96%	537	WP_016639401.1
4525	ROK family transcriptional regulator	*S. aurantiacus*	98%	414	WP_037657816.1
4526	Hypo				
4527	Hypo				
4528	plasmid stabilization protein	*S. nanshensis*	88%	108	WP_070202691.1
4529	Polyketide cyclase / dehydrase and lipid transport	*Streptomyces* sp. 1222.5	80%	137	SED31948.1

4530	Low				
4531	arsenite methyltransferase Methyltransferase domain-containing protein	*S. aurantiacus* *Streptomyces* sp. cf386	97% 82%	269	WP_016639390.1 SDP52034.1
4532	N-acetylmuramoyl-L-alanine amidase amidase	*S. aurantiacus* *S. niveus* NCIMB 11891	98% 70%	667	WP_052028977.1 EST18594.1
4533	Hypo				
4534	Hypo				
4535	Leucine Rich Repeat	*S. indicus*	67%	241	SDK18001.1
4536	putative L-ornithine 5-monooxygenase L-lysine 6-monooxygenase	*S. aurantiacus* *S. aureocirculatus*	99% 88%	467	WP_016639384.1 WP_030559018.1
4537	bifunctional polymyxin resistance protein ArnA methionyl-tRNA formyltransferase	*S. aurantiacus* *S. scabiei*	98% 85%	315	WP_016639383.1 WP_013005934.1
4538	acetylglutamate kinase	*S. aurantiacus*	99%	316	WP_016639382.1
4539	amino acid acetyltransferase	*Streptomyces* sp. RTd22	78%	181	WP_063731535.1
4540	putative Enterochelin esterase	*S. aurantiacus* JA 4570	98%	367	EPH45668.1
4541	ABC transporter ATP-binding protein putative Lipid A export ATP-binding/permease protein MsbA	*S. aurantiacus* *S. aurantiacus* JA 4570	99% 99%	566	WP_037657812.1 EPH45667.1
4542	non-ribosomal peptide synthetase non-ribosomal peptide synthase domain TIGR01720/amino acid adenylation domain-containing protein	*S. aurantiacus* *S. melanosporofaciens*	97% 85%	3680	WP_016639378.1 SED22675.1
4543	ABC transporter ATP-binding protein	*S. aurantiacus*	99%	595	WP_037657810.1
4544	iron siderophore-binding protein	*S. aurantiacus*	98%	350	WP_037657808.1
4545	ABC transporter ATP-binding protein	*S. aurantiacus*	99%	287	WP_016639375.1
4546	iron ABC transporter permease Fe(3+) enterobactin transport system permease	*S. aurantiacus* *S.* azureus	99% 86%	371	WP_037657806.1 GAP50642.1
4547	iron ABC transporter permease	*S. aurantiacus*	99%	351	WP_078620990.1
4548	bifunctional polymyxin resistance protein ArnA methionyl-tRNA formyltransferase	*S. aurantiacus* *S. melanosporofaciens*	100% 92%	315	WP_016639371.1 SED22532.1
4549	MbtH family protein	*S. aurantiacus*	100%	69	WP_016639372.1
4550	UMF1 family MFS transporter	*Bradyrhizobium* sp. DFCI-1	41%	103	ERF80733.1
4551	Hypo				
4552	amidohydrolase	*S. aurantiacus*	96%	528	WP_016640642.1
4553	DUF397 domain-containing protein	*S. aurantiacus*	96%	81	WP_016640641.1
4554	acetyl-CoA acetyltransferase	*Streptomyces* sp. NRRL S-920	95%	395	WP_030775754.1
4555	lipid-transfer protein Acetyl-CoA acetyltransferase	*Streptomyces* sp. WM6378 *Streptomyces* sp. 3213	87% 88%	352	WP_053728662.1 SEE93383.1
4556	DNA-binding protein	*S. aureocirculatus*	88%	318	WP_030559028.1
4557	enoyl-CoA hydratase	*S. aurantiacus*	99%	276	WP_016640637.1
4558	acyl-CoA synthetase	*S. aurantiacus*	99%	537	WP_016640636.1
4559	molybdenum cofactor biosysynthesis protein	*Cryptosporangium arvum*	61%	190	WP_035859205.1
4560	oxidoreductase	*S. aurantiacus*	95%	314	WP_016640634.1
4561	alpha/beta hydrolase	*S. aurantiacus*	99%	235	WP_016640633.1
4562	chorismate mutase	*S. aurantiacus*	97%	120	WP_016640632.1
4563	phenylacetate-CoA ligase	*S. aurantiacus*	97%	431	WP_016640628.1
4564	acyl-CoA synthetase	*S. jeddahensis*	87%	510	WP_067279451.1
4565	penicillin amidase	*S. aureocirculatus*	57%	240	WP_030559038.1
4566	Hypo				
4567	non-ribosomal peptide synthetase	*S. aurantiacus*	99%	3595	WP_016640626.1
4568	non-ribosomal peptide synthetase	*S. aurantiacus*	97%	2719	WP_016640624.1
4569	polyketide synthase	*Nocardia anaemiae*	54%	1591	WP_062982512.1
4570	putative Dimodular nonribosomal peptide synthase	*S. aurantiacus*	98%	637	WP_016640622.1
4571	3-oxoacyl-ACP synthase III	*S. aurantiacus*	99%	340	WP_016640621.1
4572	MFS transporter	*S. rimosus*	91%	492	WP_030670005.1
4573	ABC transporter ATP-binding protein	*S. aureocirculatus*	88%	327	WP_030574084.1

4574	N-acyl homoserine lactonase family protein	*S. aurantiacus*	98%	343	WP_016640617.1
4575	ribonucleoside-diphosphate reductase subunit alpha	*Streptomyces* sp. NRRL S-1813	84%	107	WP_030983118.1
4576	acyl-CoA synthetase	*S. mirabilis*	88%	517	WP_075025815.1
4577	nitronate monooxygenase NAD(P)H-dependent flavin oxidoreductase YrpB, nitropropane dioxygenase family	*S. aureocirculatus* *Streptomyces* sp. KS_5	92% 88%	372	WP_030559041.1 SEE13106.1
4578	putative amidohydrolase YhaA	*S. aurantiacus*	98%	421	WP_016642185.1
4579	arylesterase GDSL family lipase	*S. aurantiacus* *S.* niveus	95% 72%	2392	WP_016642186.1 WP_078073736.1
4580	TetR/AcrR family transcriptional regulator	*S. aurantiacus*	92%	226	WP_016642187.1
4581	NAD(P)/FAD-dependent oxidoreductase FAD-dependent oxidoreductase	*S. aurantiacus* *S. silaceus*	98% 80%	519	WP_016642189.1 WP_076682780.1
4582	serine/threonine protein kinase Tetratricopeptide repeat-containing protein	*S. silvensis* *Streptomyces* sp. cf386	89% 83%	761	WP_058847599.1 SDN46250.1
4583	DUF2236 domain-containing protein	*S. aurantiacus*	98%	312	WP_016642192.1
4584	Hypo				
4585	Hypo				
4586	HEXXH motif domain-containing protein	*S. afghaniensis*	66%	484	WP_020276194.1
4587	Low				
4588	cysteine hydrolase Nicotinamidase-related amidase	*S. aurantiacus* *Saccharopolyspora hirsuta*	96% 80%	207	WP_016642193.1 SEG96388.1
4589	AraC family transcriptional regulator	*S. aurantiacus*	98%	330	WP_016642194.1
4590	MFS transporter	*S. azureus*	93%	536	WP_059422516.1
4591	isovaleryl-CoA dehydrogenase	*S. aurantiacus*	98%	396	WP_016642196.1
4592	DUF1275 family protein	*S. aurantiacus*	96%	242	WP_016642197.1
4593	Predicted lipoprotein with conserved Yx(FWY)xxD motif	*Streptomyces* sp. cf386 Cov 83	68%	320	SDM73694.1
	Cluster 31 T3PKS				
4828	alpha-ketoglutarate-dependent taurine dioxygenase	*S. aurantiacus*	97%	326	WP_016642519.1
4829	dihydrolipoyl dehydrogenase	*S. aurantiacus*	99%	467	WP_016642520.1
4830	putative membrane protein	*S. aurantiacus*	98%	727	WP_016642522.1
4831	MFS transporter	*S. aureocirculatus*	81%	436	WP_030565852.1
4832	isocitrate lyase/phosphoenolpyruvate mutase family protein PEP phosphonomutase	*S. aurantiacus* *S. silvensis*	97% 85%	265	WP_016643005.1 WP_078619485.1
4833	Hypo				
4834	class F sortase Sortase (surface protein transpeptidase)	*S. aurantiacus* *Streptomyces* sp. Ncost-T10-10d	99% 65%	223	WP_016645723.1 SCF70742.1
4835	N-acetylmuramoyl-L-alanine amidase	*S. aureocirculatus*	89%	271	WP_051852838.1
4836	Hypo				
4837	RNA polymerase subunit sigma-24	*S. aureocirculatus*	90%	412	WP_030565863.1
4838	LysR family transcriptional regulator	*S. aurantiacus*	98%	356	WP_016645727.1
4839	Hypo				
4840	asparagine synthetase B asparagine synthase (glutamine-hydrolyzing)	*S. aurantiacus* *S. silvensis*	99% 91%	613	WP_016645729.1 WP_058848776.1
4841	TetR family transcriptional regulator	*S. aurantiacus*	97%	214	WP_016645730.1
4842	EamA family transporter	*S. silvensis*	82%	323	WP_058848918.1
4843	TetR family transcriptional regulator	*S. aurantiacus*	98%	198	WP_016638255.1
4844	gfo/Idh/MocA family oxidoreductase	*S. aurantiacus*	99%	314	WP_037656822.1
4845	transcriptional regulator	*S. aurantiacus*	99%	273	WP_016638253.1
4846	polyketide synthase PhlD	*S. aurantiacus* *S. kanamyceticus*	99% 78%	346	WP_016638252.1 WP_055546752.1
4847	Hypo				
4848	membrane protein	*S. aureocirculatus*	95%	222	WP_037685937.1

4849	N-acetyltransferase	*S. aurantiacus*	97%	261	WP_037656820.1
4850	Nicotinamidase-related amidase	*S. melanosporofaciens*	74%	184	SEC79059.1
4851	MFS transporter	*S. kanamyceticus*	84%	489	WP_055546758.1
4852	MarR family transcriptional regulator	*Streptomyces* sp. NRRL B-1347	80%	149	WP_030676587.1
4853	isoaspartyl peptidase/L-asparaginase	*S. aurantiacus*	94%	389	WP_016638248.1
4854	Low				
4855	acyl-CoA dehydrogenase	*S. aurantiacus*	99%	408	WP_016638246.1
4856	TetR family transcriptional regulator	*S. aurantiacus*	100%	218	WP_016638245.1
4857	peptide deformylase	*S. aurantiacus*	99%	181	WP_016638244.1
4858	UDP-N-acetylmuramoylalanine--D-glutamate ligase	*S. aurantiacus*	99%	412	WP_016638243.1
4859	putative Cobyric acid synthase glutamine amidotransferase	*S. aurantiacus* JA 4570 *S. aureocirculatus*	99% 97%	245	EPH46834.1 WP_030565884.1
4860	6-phosphofructokinase	*S. aurantiacus*	99%	341	WP_016638241.1
4861	cytochrome c oxidase assembly protein	*S. aurantiacus*	100%	319	WP_037656817.1
4862	integral membrane protein	*Streptomyces* sp. SPB074	96%	52	EDY45690.1
4863	protease inhibitor SIL-V5	*S. aurantiacus*	95%	132	WP_016638238.1
4864	1-acyl-sn-glycerol-3-phosphate acyltransferase	*S. aurantiacus*	99%	147	WP_016638237.1
4865	Ala-tRNA(Pro) deacylase	*S. indicus*	84%	161	SDK19708.1
4866	alpha/beta hydrolase	*S. aurantiacus*	99%	538	WP_016638235.1
4867	NERD nuclease	*Streptomyces* sp. PRh5	83%	134	EXU63185.1
4868	Hypo				
	Cluster 32 Ectoine				
5596	amidohydrolase	*S. aureocirculatus*	91%	363	WP_030571725.1
5597	alanine--glyoxylate aminotransferase family protein aspartate aminotransferase	*S. aurantiacus* *S. kanamyceticus*	99% 92%	366	WP_016644018.1 WP_079044006.1
5598	diaminobutyrate acetyltransferase	*S. aurantiacus*	98%	167	WP_016644019.1
5599	diaminobutyrate--2-oxoglutarate transaminase aspartate aminotransferase family protein	*S. aureocirculatus* *S. silvensis*	95% 95%	423	WP_030571735.1
5600	L-ectoine synthase	*S. aurantiacus*	99%	131	WP_016643612.1
5601	ectoine hydroxylase	*S. aurantiacus*	99%	299	WP_016643610.1
5602	putative 12-oxophytodienoate reductase 1 alkene reductase	*S. aurantiacus* JA 4570 *S. aurantiacus*	99% 99%	356	EPH41450.1 WP_037662225.1
5603	MarR family transcriptional regulator	*S. aurantiacus*	99%	178	WP_016643608.1
5604	cysteine desulfurase class V aminotransferase	*S. aurantiacus* *S. silvensis*	95% 87%	360	WP_016643607.1 WP_079086937.1
5605	DsbA family oxidoreductase	*S. aurantiacus*	99%	243	WP_016643606.1
	Cluster 33 NRPS				
6226	LysR family transcriptional regulator	*S. aurantiacus*	95%	301	WP_016642589.1
6227	pyroglutamyl peptidase	*S. aurantiacus*	88%	424	WP_052029188.1
6228	membrane protein CysZ protein	*S. iakyrus* *Streptomyces* sp. 1222.5	79% 76%	266	WP_033306134.1 SEC84700.1
6229	peptidase M28 zinc carboxypeptidase	*S. aureocirculatus* *S.* ipomoeae	86% 72%	857	WP_030566353.1 WP_009297172.1
6230	helix-turn-helix transcriptional regulator	*S. aurantiacus*	98%	339	WP_037661070.1
6231	organic hydroperoxide resistance protein	*S. aurantiacus*	99%	142	WP_016642594.1
6232	NADP-dependent oxidoreductase	*S. aurantiacus*	99%	344	WP_016642595.1
6233	MarR family transcriptional regulator	*S. silvensis*	89%	160	WP_058845882.1
6234	Hypo				
6235	PucR C-terminal helix-turn-helix domain-containing protein	*Streptomyces* sp. KS_5	81%	446	SED45806.1
6236	Hypo				
6237	DUF3068 domain-containing protein	*S. aurantiacus*	97%	344	WP_037661073.1
6238	D-inositol-3-phosphate glycosyltransferase	*S. aurantiacus*	99%	391	WP_016642601.1
6239	SAM-dependent methyltransferase	*S. aurantiacus*	99%	239	WP_016642602.1

6240	putative Alpha-(1->3)-arabinofuranosyltransferase	*S. aurantiacus* JA 4570	98%	1580	EPH42486.1
6241	condensation protein	*S. silvensis*	86%	465	WP_058845886.1
6242	class I SAM-dependent methyltransferase 2-polyprenyl-3-methyl-5-hydroxy-6-metoxy-1,4-benzoquinol methylase	*S. aurantiacus* *S. indicus*	96% 82%	254	WP_016642605.1 SDK58915.1
6243	Hypo				
6244	Hypo				
6245	methyltransferase FkbM	*S. silvensis*	84%	274	WP_058845888.1
6246	Hypo				
6247	serine/threonine protein kinase	*S. ruber*	69%	590	WP_030358572.1
6248	acyltransferase	*S. aurantiacus*	95%	433	WP_016642611.1
6249	histidine kinase	*S. kanamyceticus*	79%	799	WP_079043755.1
6250	malonate--CoA ligase	*S. aurantiacus*	98%	491	WP_016642613.1
6251	Hypo				
6252	Low				
6253	peptidase S8	*Streptomyces* sp. NRRL B-1347	76%	1250	WP_078869234.1
6254	Hypo				
6255	serine protease	*S. aurantiacus*	96%	586	WP_037661089.1
6256	MoxR family ATPase	*Streptomyces* sp. NRRL B-1347	88%	347	WP_030684826.1
6257	Low				
6258	MFS transporter	*S. aurantiacus*	98%	477	WP_016642619.1
6259	rod shape-determining protein	*S. aurantiacus*	100%	345	WP_016642620.1
6260	histidine kinase	*S. aurantiacus*	97%	492	WP_016642621.1
6261	molybdenum ABC transporter permease subunit	*S. niveus*	79%	654	WP_078077475.1
6262	molybdate-binding protein	*S. aurantiacus*	96%	277	WP_016638397.1
	Cluster 34 Nucleoside				
6275	type I methionyl aminopeptidase	*S. aurantiacus*	96%	261	WP_016638384.1
6276	sulfate adenylyltransferase	*S. aurantiacus*	97%	432	WP_016638383.1
6277	sulfate adenylyltransferase	*S. aurantiacus*	98%	299	WP_037656932.1
6278	L-ornithine 5-monooxygenase oxidoreductase	*S. aurantiacus* *S. hygroscopicus*	99% 64%	442	WP_016638381.1 WP_014670019.1
6279	putative hydroxylase	*S. cacaoi* subsp. asoensis	53%	216	ABX24500.1
6280	sulfotransferase family protein	*S. aurantiacus*	96%	299	WP_037656930.1
6281	putative hydroxylase	*S. cacaoi* subsp. *asoensis*	35%	208	ABX24501.1
6282	putative Argininosuccinate lyase 2	*S. aurantiacus*	96%	444	WP_016638377.1
6283	putative UDP-N-acetylglucosamine 1-carboxyvinyltransferase	*S. aurantiacus* JA 4570	99%	449	EPH46689.1
6284	protein-tyrosine-phosphatase	*Streptomyces* sp. NRRL S-350 Cov 80	36%	223	WP_030247790.1
6285	radical SAM protein	*S. aurantiacus*	99%	436	WP_016638373.1
6286	putative Histidinol-phosphate aminotransferase	*S. aurantiacus* JA 4570	98%	384	EPH46685.1
6287	Low				
6288	2OG-Fe(II) oxygenase	*Methylotenera versatilis*	35%	207	WP_013147946.1
6289	putative L-lysine 2,3-aminomutase	*S. aurantiacus*	97%	462	WP_016638369.1
6290	pyridoxal phosphate-dependent aminotransferase	*S. aurantiacus*	97%	432	WP_016638368.1
6291	hypo				
6292	ATP-dependent carboxylate-amine ligase domain protein nikkomycin biosynthesis protein SanS	*S. acidiscabies* *S.* ansochromogenes	47% 45%	405	WP_010359211.1 AAK53061.1
6293	2OG-Fe(II) oxygenase superfamily protein	*Actinoplanes derwentensis*	61%	264	SDS41503.1
6294	GrpB domain, predicted nucleotidyltransferase, UPF0157 family	*Streptomyces*	70%	203	SCK09282.1
6295	MFS transporter	*S. aurantiacus*	98%	432	WP_016638363.1

	Predicted arabinose efflux permease, MFS family	*Micromonospora echinofusca*	50%		SCG16244.1
	Cluster 35 Siderophore				
6629	DUF218 domain-containing protein	*S. clavuligerus* ATCC 27064	80%	223	EFG07774.1
6630	ABC transporter substrate-binding protein	*S. aureocirculatus*	88%	345	WP_030563404.1
6631	L-2,4-diaminobutyrate decarboxylase putative amino acid decarboxylase, pyridoxal-dependent protein	*S. aurantiacus* *S. ipomoeae*	98% 82%	480	WP_016639537.1 WP_009312397.1
6632	alcaligin biosynthesis protein L-lysine 6-monooxygenase (NADPH)	*S. aurantiacus* *Streptomyces* sp. ACT-1	97% 81%	427	WP_016639538.1 WP_003968994.1
6633	putative Rhizobactin siderophore biosynthesis protein RhbD acetyltransferase	*S. aurantiacus* *S. silvensis*	99% 86%	205	WP_016639539.1 WP_058848519.1
6634	IucA/IucC family siderophore biosynthesis protein	*S. aurantiacus*	98%	591	WP_016639540.1
6635	beta-N-acetylhexosaminidase	*S. aurantiacus*	95%	554	WP_052028982.1
6636	Tat pathway signal sequence domain protein	*S. silvensis*	88%	295	WP_058848516.1
6637	Hypo				
6638	glutamine-fructose-6-phosphate transaminase (isomerizing)	*S. aurantiacus*	100%	605	WP_016639544.1
6639	Hypo				
	Cluster 36 Melanin				
6788	Hypo				
6789	IclR family transcriptional regulator	*S. silvensis*	99%	213	WP_058848405.1
6790	ATP-dependent helicase	*S. aurantiacus*	99%	597	WP_016644284.1
6791	MFS transporter	*S. aurantiacus*	98%	488	WP_016644283.1
6792	Tyrosinase cofactor	*S. aurantiacus*	96%	131	WP_016644282.1
6793	tyrosinase MelC2	*S. aurantiacus* *S.* tanashiensis	97% 82%	275	WP_016644281.1 BAB20029.1
6794	cysteine hydrolase Nicotinamidase-related amidase	*Actinobacteria bacterium* OK006 *Verrucosispora sediminis*	77% 77%	190	WP_054236768.1 SFC28147.1
6795	phosphotransferase	*Streptomyces* sp. NRRL S-920	78%	182	WP_078594900.1
6796	xanthine dehydrogenase subunit D oxidoreductase	*S. aurantiacus* *S. aureocirculatus*	99% 96%	765	WP_016644278.1 WP_030568722.1
6797	iron-suplhur oxidoreductase (2Fe-2S)-binding protein	*S.* griseo*aurantiacus* Cov 54 *Actinobacteria bacterium* OK006 Cov 50	68% 75%	907	WP_006139389.1 WP_054229665.1
	Cluster 37 Other				
7187	acyl-CoA dehydrogenase	*S. aurantiacus*	92%	309	WP_037664330.1
7188	medium-chain specific acyl-CoA dehydrogenase	*S. aurantiacus*	99%	380	WP_016645418.1
7189	Triostin synthetase I cyclohex-1-ene-1-carboxylate:CoA ligase	*S. aurantiacus* *S. aureocirculatus*	97% 91%	517	WP_016645419.1 WP_030569402.1
7190	calcium-binding protein	*S. aureocirculatus*	96%	165	WP_030569399.1
7191	anti-sigma factor antagonist	*S. aurantiacus*	99%	124	WP_016645421.1
7192	RNA polymerase sigma factor SigK	*S. aurantiacus*	100%	198	WP_016645422.1
7193	MDMPI N domain containing protein	*S. aurantiacus*	94%	530	WP_078621732.1
7194	formyltetrahydrofolate deformylase	*S. aurantiacus*	100%	297	WP_016641185.1
7195	Hypo				
7196	lipoprotein	*S. aurantiacus*	96%	470	WP_037659548.1
7197	Hypo				
7198	transcriptional regulator	*Streptomyces* sp.	92%	462	WP_085211592.1

		Amel2xC10			
7199					
7200	ABC transporter ATP-binding protein	*S. aurantiacus*	98%	257	WP_037659560.1
7201	transporter	*S. aurantiacus*	77%	544	WP_079103965.1
7202	SAM-dependent methyltransferase	*Nocardia pseudobrasiliensis*	66%	408	WP_067999301.1
7203	Thioester reductase	*Allokutzneria albata*	61%	1053	WP_081900362.1
7204	putative Desosaminyl transferase EryCIII	*S. aurantiacus*	97%	444	WP_016641176.1
7205	class I SAM-dependent methyltransferase NbmE	*S. aurantiacus* *S. narbonensis*	97% 55%	258	WP_016641175.1 AAM88354.1
7206	P450-derived glycosyltransferase activator	*S. resistomycificus*	35%	362	WP_030038566.1
7207	putative Desosaminyl transferase EryCIII	*S. aurantiacus* JA 4570	97%	418	EPH43893.1
7208	putative Glucose-1-phosphate thymidylyltransferase 1	*S. aurantiacus* JA 4570	98%	287	EPH43892.1
7209	Pyridoxamine 5'-phosphate oxidase	*Actinopolyspora xinjiangensis*	55%	183	SDP37014.1
7210	NDP-hexose 2,3-dehydratase	*Actinosynnema* sp. ALI-1.44	56%	485	WP_076991945.1
7211	WxcM-like, C-terminal	*Streptomyces* sp. Cmuel-A718b	72%	163	SCF70151.1
7212	DegT/DnrJ/EryC1/StrS family aminotransferase	*S. aurantiacus*	98%	369	WP_016641168.1
7213	dTDP-glucose 4,6-dehydratase	*S. aurantiacus*	94%	337	WP_037659538.1
7214	putative UDP-glucose 4-epimerase NAD-dependent epimerase/dehydratase	*S. aurantiacus* JA 4570 *S. xiamenensis*	94% 64%	348	EPH43886.1
7215	putative dTDP-4-dehydrorhamnose 3,5-epimerase	*S. aurantiacus*	99%	216	WP_016641165.1
7216	ABC transporter permease	*Rhodococcus erythropolis*	52%	251	WP_021332328.1
7217	daunorubicin resistance ATP-binding protein	*S. aurantiacus*	99%	318	WP_016641163.1
7218	TetR family transcriptional regulator	*S. aurantiacus*	98%	253	WP_037659536.1
7219	putative Regulator protein	*S. aurantiacus* JA 4570	99%	623	EPH43881.1
7220	diguanylate cyclase	*S. silvensis*	91%	469	WP_058846210.1
7221	ABC transporter permease	*S. aurantiacus*	98%	402	WP_016641779.1
7222	ABC transporter ATP-binding protein	*S. aurantiacus*	98%	262	WP_016641778.1